Parameter estimation via differential algebra and operational culculus

Mamadou Mboup

Parameter estimation via differential algebra and operational calculus

Mamadou MBOUP

UFR de mathématiques et informatiques, Université Paris Descartes - CRIP5
45, rue des Saints-Pères, 75270 Paris cedex 06 - France
&
ALIEN, INRIA-Futurs

Abstract

Parameter estimation is approached via a new standpoint, based on differential algebra and operational calculus. Some appications such as, the estimation of a noisy damped sinusoid, the analysis of chirp signal, the detection of piecewise polynomial signals and their discontinuities are presented with numerical simulations.

Key words: Parametric estimation, Differential algebra, Operational calculus, Continuous-time signals, Jacobi orthogonal polynomials.

1 Introduction

Parametric estimation may often be formalized as follows

$$y = F(x, \boldsymbol{\Theta}) + n \tag{1}$$

where

- the observed signal y is a functional F of the "true" signal x, which depends on a set $\boldsymbol{\Theta}$ of parameters,
- n is a noise corrupting the observation.

Finding a "good" approximation of the components of $\boldsymbol{\Theta}$ is the subject of a huge literature in various fields of applied mathematics.

Email address: `Mamadou.Mboup@math-info.univ-paris5.fr` (Mamadou MBOUP).

This problem is approached here with a new standpoint (see also [1] and [2]) which is based on the following tools, which are of algebraic flavour:

- Differential algebra [3,4], which plays with respect to differential equations a similar rôle to commutative algebra with respect to algebraic equations. It was introduced in nonlinear control theory by M. Fliess [5] almost twenty years ago for understanding some specific questions like input-output inversion [6], [7]. It has allowed to recast the whole of nonlinear control into a more realistic light. The best example is of course the discovery of *flat* systems [8] which prove useful in practice (see *e.g.* [9]).
- Operational calculus [10], [11], [12] which was a most classical tool among control and mechanical engineers. Operational calculus is often formalised via the Laplace transform whereas the Fourier transform is today the cornerstone in estimation [1].

The estimation method is presented in section 2. We begin with two simple introductory examples in order to fix the notations and to describe the main steps of the corresponding estimation algorithm. Section 3 is devoted to the mathematical background. Therein, the basic notions of differential algebra and operational calculus are reviewed with the aim of keeping the presentation in a tutorial level. Various type of identifiability are defined, some of them being often encountered in practice. Most parametric estimation methods are devised upon the orthogonality principle. We show in section 4 that the presented method does not depart from this rule. In particular, the noise influence is discussed and some links with the popular least squares approach are pointed out. Parameter estimation in nonstationary context is illustrated further in section 5, through two application examples. The first one concerns the estimation of a chirp signal. The estimates of the chirp parameters are obtained on a short segment of a (coloured) noisy observation signal. The second application deals with the change point detection problem (see [13]), recasted here into delay estimation. A closed form expression of the delay, representing the change point location, is provided. Finally, section 6 is devoted to the concluding remarks.

[1] Note that the one-sided Laplace transform is causal, but not the Fourier transform over $\mathbb{R}$.

2 Examples and algorithm

2.1 Complex damped sinusoidal signal

Consider the estimation of the parameters α and ω of the complex damped sinusoidal signal

$$y(t) = ae^{(\alpha+i\omega)t} + \gamma + n(t), \quad t \geqslant 0 \tag{2}$$

with a constant bias perturbation γ and an additive noise corruption $n(t)$. Such an estimation problem arises in many data analysis applications. It has a long and rich history [14],[15], [16], [17], [18], [19] because 1) frequency estimation [2] is fundamental in signal processing and 2) the signal is transient due to the damping factor and this makes the problem challenging.

As quoted in the introduction, the estimation will be based on a short time-segment of the observed signal. Now, note that in any finite interval time $I_\tau^T = [\tau,\, \tau + T]$, the noise $n(t)$ may be decomposed as

$$n(t) = \gamma_\tau^T + n_\tau^T(t), \tag{3}$$

i.e. the sum of a constant γ_τ^T, representing its mean (average) value and a zero-mean term, $n_\tau^T(t)$. Therefore, we subsequently interpret the overall perturbation $\gamma + n(t)$ in (2) according to the above decomposition and we consider, without any loss of generality, that $n(t)$ is zero-mean.

2.1.1 Differential equation

Let

$$x(t) = ae^{(\alpha+i\omega)t} + \gamma \tag{4}$$

denotes the unobserved *structured* part of $y(t)$. The starting point of our estimation method is to observe that $x(t)$ satisfies a linear differential equation

$$\dot{x}(t) = (\alpha + i\omega)(x(t) - \gamma), \tag{5}$$

where the unknown parameters α and ω appear explicitly. Translated into the operational domain (see section 3.2), this differential equation reads as:

$$s\hat{x}(s) = (\alpha + i\omega)\hat{x}(s) + x(0) - (\alpha + i\omega)\frac{\gamma}{s}. \tag{6}$$

[2] see also [20] for an algebraic frequency estimation.

2.1.2 Algorithm

The main steps towards the estimation of $\theta \triangleq \alpha + i\omega$ are now described in the next algorithm, where for notational convenience, we write $\hat{x}$ and $\hat{x}'$ for $\hat{x}(s)$ and $\frac{d}{ds}\hat{x}(s)$, respectively.

Step 1 - Elimination of structured perturbations Since we are not concerned with the estimation of the amplitude a, the initial condition $x(0)$ is considered as an undesired perturbation, like the constant bias γ. Note that these perturbations are easily anihilated by multiplying both sides of (6) by s, followed by a derivation of order 2, with respect to s. This amounts to applying the linear differential operator

$$\Pi = \frac{d^2}{ds^2} s = s\frac{d^2}{ds^2} + 2\frac{d}{ds}$$

to both members of (6). The resulting equation, given by

$$s^2\hat{x}'' + 4s\hat{x}' + 2\hat{x} = (s\hat{x}'' + 2\hat{x}')\theta, \tag{7}$$

shows that the constant bias γ and the unknown initial condition $x(0)$ will have no effect on the estimation result.

Step 2 - Linear estimator Replace x in (7) by the observed signal y. We thus obtain a linear estimator $\widetilde{\theta}$ for θ, given by:

$$s^2\hat{y}'' + 4s\hat{y}' + 2\hat{y} = (s\hat{y}'' + 2\hat{y}')\widetilde{\theta}, \tag{8}$$

Step 3 - (Strictly) proper estimator Recall that derivation with respect to s in the operational domain translates into multiplication by $-t$ in the time domain. Multiplication by s in the operational domain, in turn, corresponds to derivation in the time domain. Implementing the linear estimator (8) in its present form is therefore not convient: derivation amplifies the high frequency components and consequently, the noise contribution. A simple solution is readily obtained by making the estimator (strictly) proper (see definition 3.4.4). For this, it suffices to multiply both sides of (8) by $s^{-\nu}$ where ν is an integer greater than or equal to the highest power of s in (8). The following estimator,

$$\left(\frac{\hat{y}''}{s^{\nu-1}} + 2\frac{\hat{y}'}{s^{\nu}}\right)\widetilde{\theta} = \frac{\hat{y}''}{s^{\nu-2}} + 4\frac{\hat{y}'}{s^{\nu-1}} + 2\frac{\hat{y}}{s^{\nu}} \tag{9}$$

is strictly proper for $\nu \geqslant 3$: only integral operators of the observed signal ($s^{-k}\hat{y}$, with $k > 0$) are involved.

Step 4 - Numerical estimate (time domain) The final step is to express the linear estimator (9) back in the time domain. To proceed, let us recall once again that

- $\frac{d}{ds} \mapsto$ "multiplication by $-t$",
- for $k > 0$, $s^{-k} \mapsto$ "iterated integral of order k". For example, if $\hat{u} \mapsto u(t)$, then $s^{-k}\hat{u}$ will correspond in the time domain to the order k iterated integral,

$$\int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_1} u(\tau) d\tau dt_1 \cdots dt_{k-1} \quad = \quad \frac{1}{(k-1)!} \int_0^t (t \; - \; \tau)^{k-1} u(\tau) d\tau, \quad (10)$$

  which reduces to a single itegral by use of the Cauchy formula for repeated integration.

Applying these rules to (9), we readily obtain the following explicit formula for an estimate $\widetilde{\theta}$ of θ, as a function of the estimation time (the estimation interval is I_0^t):

$$\widetilde{\theta}_t = \frac{\displaystyle\int_0^t p_{\nu,t}(\tau)(t-\tau)^{\nu-3} y(\tau) d\tau}{\displaystyle\int_0^t \{(\nu+1)\tau - 2t\}(t-\tau)^{\nu-2} \tau y(\tau) d\tau}, \quad (11)$$

$$= \widetilde{\alpha}_t + i\widetilde{\omega}_t. \quad (12)$$

where $p_{\nu,t}(\tau) = (\nu-1)(\nu-2)\tau^2 - 4(\nu-1)(t-\tau)\tau + 2(t-\tau)^2$. Let us quote the following remarks.

- The estimation time t may be small, resulting in fast estimation.
- The noise effect is attenuated by the iterated integrals, which behave as low pass filtering.
- The computational complexity is low.

2.2 Real damped sinusoidal signal

We proceed with the same estimation problem but here we consider that only the imaginary part of the signal (2),

$$y(t) = ae^{\alpha t} \sin(\omega t) + \gamma + n(t), \quad t \geqslant 0 \quad (13)$$

is observed.

As before, let $x(t) \triangleq y(t) - n(t)$ denotes the unobserved structured part of $y(t)$. Then one may easily check that it satisfies the linear differential equation

$$\ddot{x}(t) = 2\alpha\dot{x}(t) - (\alpha^2 + \omega^2)(x(t) - \gamma), \quad (14)$$

which coefficients depend on the unknown parameters. This equation is clearly nonlinear in the parameters α and ω, subsequently gathered in the vector $\mathbf{\Theta} = (\alpha, \omega)$. However, if we define $\theta'_1 \triangleq 2\alpha$ and $\theta'_2 \triangleq -(\alpha^2 + \omega^2)$, then it becomes linear in the new parameter vector

$$\mathbf{\Theta}' = \begin{bmatrix} \theta'_1 \\ \theta'_2 \end{bmatrix}.$$

Thus, θ'_1 and θ'_2 are now on considered as two independent unknown parameters, the estimation of which is the problem under concern. Translating (14) into the operational domain, we obtain, after anihilating the initial conditions and the constant bias γ,

$$\begin{aligned} s^3\hat{x}^{(3)} + 9s^2\hat{x}'' + 18s\hat{x}' + 6\hat{x} \\ = (s^2\hat{x}^{(3)} + 6s\hat{x}'' + 6\hat{x}')\theta'_1 + (s\hat{x}^{(3)} + 3\hat{x}'')\theta'_2. \end{aligned} \tag{15}$$

Remark 1 *If the amplitude a of the signal were also to be estimated, then, instead of* (14)*, one would have started with e.g. the fourth order differential equation*

$$\frac{d^4x}{dt^4} = \theta'_1\frac{d^3x}{dt^3} + \theta'_2\frac{d^2x}{dt^2}.$$

In the operational domain, one would get:

$$s^4\hat{x} = (s^3\hat{x})\theta'_1 + (s^2\hat{x})\theta'_2 + \sum_{i=0}^{3}\lambda_i s^i, \tag{16}$$

where the coefficients λ_i are linear combinations of the initial conditions $x_0^{(i)}$, $i = 0, \ldots, 3$. The first coefficient, given by $\lambda_0 = x_0^{(3)} - \theta'_1 x''_0 - \theta'_2 x'_0$, is a function of the amplitude a which does not depend on x_0 and hence on the constant biais γ. It would then have been considered as part of the parameters to be estimated and the others, as undesired perturbations. Applying the differential operator

$$\Pi = \frac{d^3}{ds^3} \cdot \frac{1}{s}$$

to (16)*, which amounts to dividing both sides of the equation by s followed by a third order differentiation with respect to s, annihilates the undesired perturbations.*

Since there are two parameters to be estimated, one has to complete (15) with a second equation in order to obtain a square system. Such a completion is now readily achieved by the derivative of both sides of equation (15) with respect to s:

$$s^3\hat{x}^{(3)} + 9s^2\hat{x}'' + 18s\hat{x}' + 6\hat{x} = (s^2\hat{x}^{(3)} + 6s\hat{x}'' + 6\hat{x}')\theta_1' + (s\hat{x}^{(3)} + 3\hat{x}'')\theta_2', \tag{17a}$$

$$s^3\hat{x}^{(4)} + 12s^2\hat{x}^{(3)} + 36s\hat{x}'' + 24\hat{x}' = (s^2\hat{x}^{(4)} + 8s\hat{x}^{(3)} + 12\hat{x}'')\theta_1' + (s\hat{x}^{(4)} + 4\hat{x}^{(3)})\theta_2' \tag{17b}$$

The functions $(s^2\hat{x}^{(3)}+6s\hat{x}''+6\hat{x}')$ and $(s\hat{x}^{(3)}+3\hat{x}'')$ being linearly independent, their Wronskian is nonzero. The resulting system is therefore (generically) invertible. Let us mention that in the operational domain, the invertibility of the system is not necessary. For example, although $\hat{x}$ and $s\hat{x}$ are proportional, their respective images in the time domain, x and $\dot{x}$, are not.

Following steps 2 and 3 above yields the estimates of θ_1' and θ_2' from the solution of

$$\widehat{P}\widetilde{\boldsymbol{\Theta}'} = \widehat{Q}, \tag{18}$$

where the matrix $\widehat{P}$ and vector $\widehat{Q}$ are given by

$$\widehat{P} = \begin{bmatrix} \frac{\hat{y}^{(3)}}{s^{\nu-2}} + 6\frac{\hat{y}''}{s^{\nu-1}} + 6\frac{\hat{y}'}{s^{\nu}} & \frac{\hat{y}^{(3)}}{s^{\nu-1}} + 3\frac{\hat{y}''}{s^{\nu}} \\ \frac{\hat{y}^{(4)}}{s^{\nu-2}} + 8\frac{\hat{y}^{(3)}}{s^{\nu-1}} + 12\frac{\hat{y}''}{s^{\nu}} & \frac{\hat{y}^{(4)}}{s^{\nu-1}} + 4\frac{\hat{y}^{(3)}}{s^{\nu}} \end{bmatrix}$$

and

$$\widehat{Q} = \begin{bmatrix} \frac{\hat{y}^{(3)}}{s^{\nu-3}} + 9\frac{\hat{y}^{(2)}}{s^{\nu-2}} + 18\frac{\hat{y}'}{s^{\nu-1}} + 6\frac{\hat{y}}{s^{\nu}} \\ \frac{\hat{y}^{(4)}}{s^{\nu-3}} + 12\frac{\hat{y}^{(3)}}{s^{\nu-2}} + 36\frac{\hat{y}''}{s^{\nu-1}} + 24\frac{\hat{y}'}{s^{\nu}} \end{bmatrix}.$$

This estimator is strictly proper for $\nu \geqslant 4$. In the time domain, we get, by step 4, an equation of the form

$$\underbrace{\left[\int_0^t P_\nu(t,\tau)y(\tau)d\tau\right]}_{\mathcal{P}_\nu(t)}\widetilde{\boldsymbol{\Theta}'}(t) = \underbrace{\int_0^t Q_\nu(t,\tau)y(\tau)d\tau}_{\mathcal{Q}_\nu(t)}, \tag{19}$$

where the entries of the 2×2-matrix $\mathcal{P}_\nu(t)$ and the 2×1-vector $\mathcal{Q}_\nu(t)$ are made up of iterated integrals up to time t of the noisy observed signal $y(t)$.

Once the estimates $\widetilde{\boldsymbol{\Theta}'}$ are obtained, the estimates $\widetilde{\boldsymbol{\Theta}}$ of the parameters $\boldsymbol{\Theta}$ are deduced from the system of algebraic equations:

$$\tilde{\alpha} = \frac{\tilde{\theta}_1'}{2} \tag{20}$$

$$\tilde{\omega}^2 = -([\tilde{\theta}_1']^2 + \tilde{\theta}_2') \tag{21}$$

The parameters $\boldsymbol{\Theta}'$, which are estimated as the solution of a linear system (19), are said to be linearly identifiable. On the other hand, the parameters $\boldsymbol{\Theta}$, which are obtained from an algebraic functions of the entries of $\boldsymbol{\Theta}'$, are said to be weakly linearly identifiable (see section 3).

2.2.1 *Numerical simulation*

Let us give some numerical examples. The next simulations show some illustrations of the behavior of the estimators given in (11) and in (19). They also provide a sketch of the overall behavior of the estimators build within our framework.

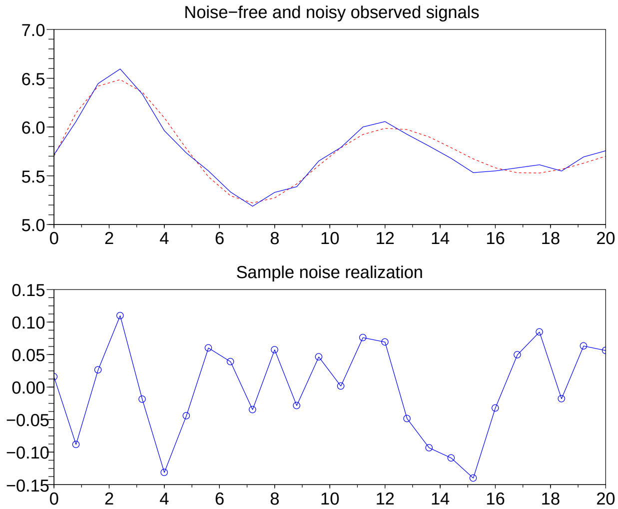
Figure 1 represents the unobserved structured part of the signal (13), $x(t)$, $t \in [0,\, 20]$ with $\alpha = -0.1, \omega = \pi/5$ and $\gamma = 5.7$ (dotted line), together with a sample realization of the noisy observed signal $y(t)$ (solid line). The subfigures illustrate three different simulation contexts that we describe below.

In the first scenario, figure 1(a), we consider a discrete-time[3] setting: $\{n(t_m)\}, t_m = mT_0$ with $m = 0, \ldots, \lfloor \frac{20}{T_0} \rfloor$ is a zero-mean white Gaussian noise, with variance σ^2. The sampling period is set to $T_0 = 0.8$, corresponding to 25 samples. If we linearly interpolate these samples, we obtain a continuous-time pseudo noise signal. Call $n_0(t)$ the pseudo noise so obtained from the above white noise, with $\sigma^2 = 1$. Likewise, let $n_k(t)$ be obtained similarly, with a different sampling period $T_k = 2^{-k}T_0$. By taking the sum of K of such noises, $n(t) = \sum_{k=0}^{K-1} A_k n_k(t)$, with $A_k = A_0^k$, for some $A_0 < 1$, one obtain a continuous-time noise. If N samples of $n(t)$ are to be simulated, then one may chose $K = \lfloor N \rfloor$ and the resulting sampling period would be $Ts = 2^{-K}T_0$. This kind of noise, so-called Perlin noise [24], is widely used in turbulence and in Computer Graphics (in its 2D or 3D version). Figure 1(b) displays a simulation using this type of noise. Finally, in the situation of figure 1(c), the noise is simulated in a manner which is, *a priori*, more in accordance with the continuous-time context than in the preceding scenarios. For this, the noisy signal $y(t), t \in [0, 20]$ is modeled by the stochastic differential equation [25]

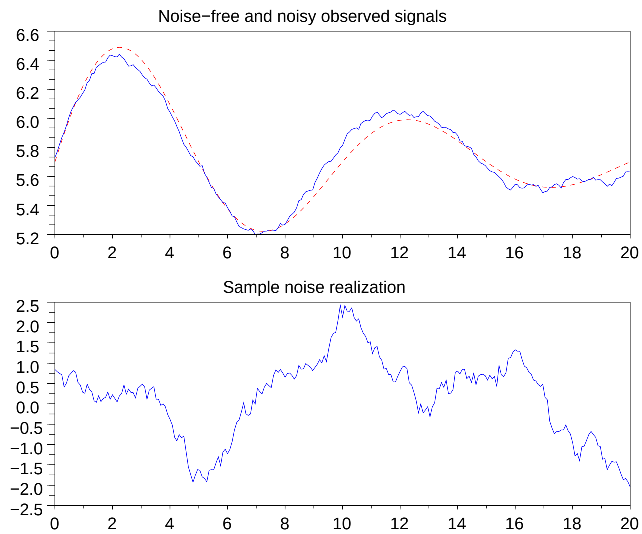
$$dy = f(t, y)dt + \sigma dB(t), \tag{22}$$

where $f(t, y) = \alpha y + \omega a e^{\alpha t} \cos(\omega t)$, $B(t)$ is a Wiener process and σ controls the level of the noise. The noisy observed signal $y(t)$ is therefore simulated as a numerical solution of (22), using the implicit Euler approximation scheme.

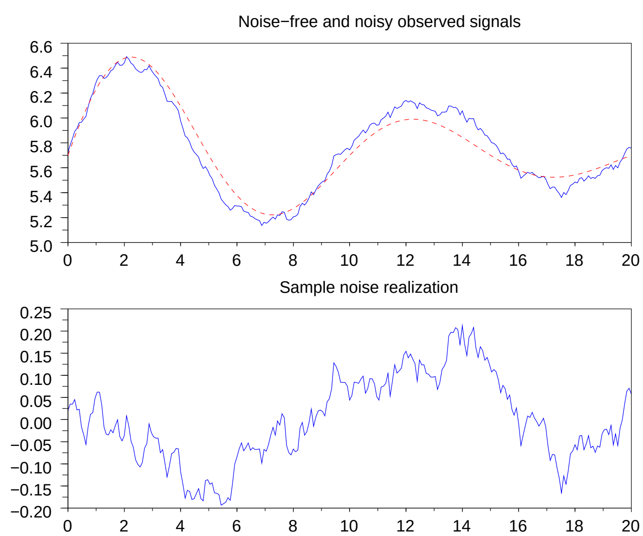
[3] Note that a fully discrete-time counterpart of the presented method can be found in [21], [22] and [23].



(a) White noise



(b) Perlin noise



(c) Continuous-time noise

Fig. 1. Signal and noise simulation: $SNR = 15dB$ in each scenario.

In all these figures, the noisy signals represent the imaginary part of (2). The noises are scaled in order to have a signal-to-noise ratio, computed as

$$SNR = 10\log_{10}\left\{\frac{\sum_{i=0}^{N-1}|y(t_i)-\gamma|^2}{\sum_{i=0}^{N-1}|n(t_i)|^2}\right\},$$

of $SNR = 15dB$. The number of simulated samples is $N = 25$ for the discrete-time context and $N = 256$ for the remaining situations.

Of these three situations, we retain only the discrete-time white noise and the Perlin noise scenarios for the next simulations. Indeed the last two situations are closely related since the Perlin's noise may be seen as a random walk, which corresponds to the infinitesimal structure of the Brownian motion.

The following simulations show the average over 100 trials of the estimates $\widetilde{\alpha}_t$ and $\widetilde{\omega}_t$, computed from (19), as a function of the estimation time $t \in [7.2, 20]$. The corresponding estimates obtained from (2) by the modified Prony's method [17] (based on nonlinear least squares fitting) are also given for comparison. Let us mention that here, and in all the subsequent simulations, the integrals are numerically computed using the "work horse" trapezoidal method.

We begin with the white noise scenario. The curves in figure 2 represent $\widetilde{\alpha}_t$

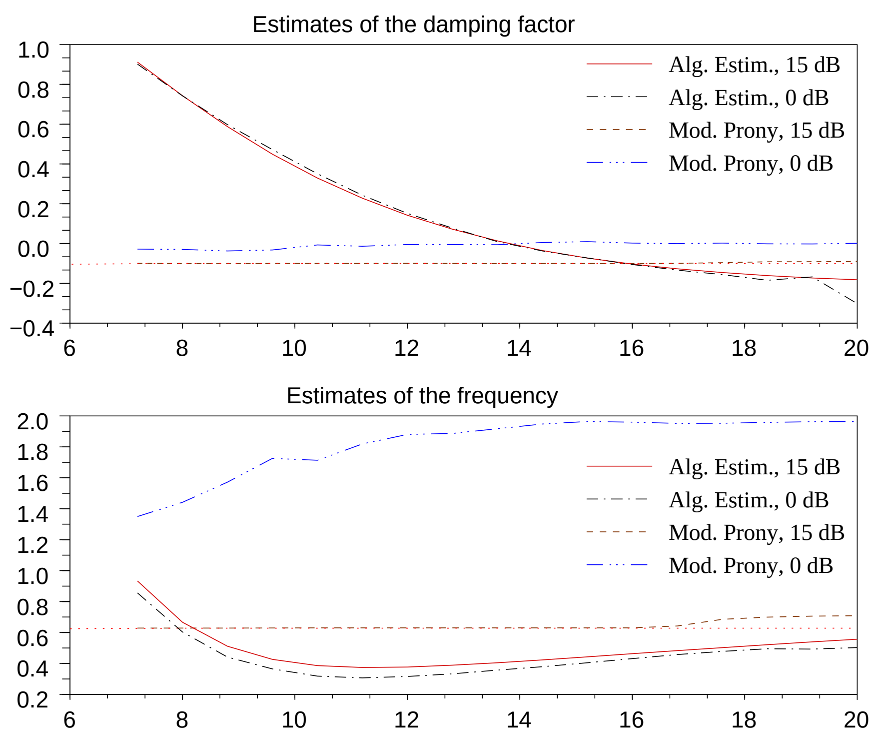


Fig. 2. Estimates of α and ω vs the estimation time.

(top graph) and $\widetilde{\omega}_t$ (bottom graph) for $SNR = 15dB$ and $SNR = 0dB$, respectively. For the modified Prony's method, we have set $\gamma = 0$ though this method is also able to handle a constant biais by directly estimating it. At moderate SNR ($15dB$), the modified Prony's method provides much better results in this scenario. As expected however, these good performance dramatically degrade as the SNR drops to $0dB$. Meanwhile, the results produced by the algebraic estimator (19) are fairly the same by changing the SNR from $15dB$ to $0dB$. This suggests that the estimation error steming from the (very high) inacuracy of the numerical integrations (10 to 25 points trapezoidal scheme) largely dominates that due to the noise.

The next experiment in figure 3, where we consider a Perlin's noise with $N = 256$, strengthens this suggestion. In order to avoid a bad behavior, due to a too small sampling period, we subsample the signals by a factor $1/8$ before applying the modified Prony's method. The constant bias, however, is no longer set to zero but to $5.7(1 + i)$.

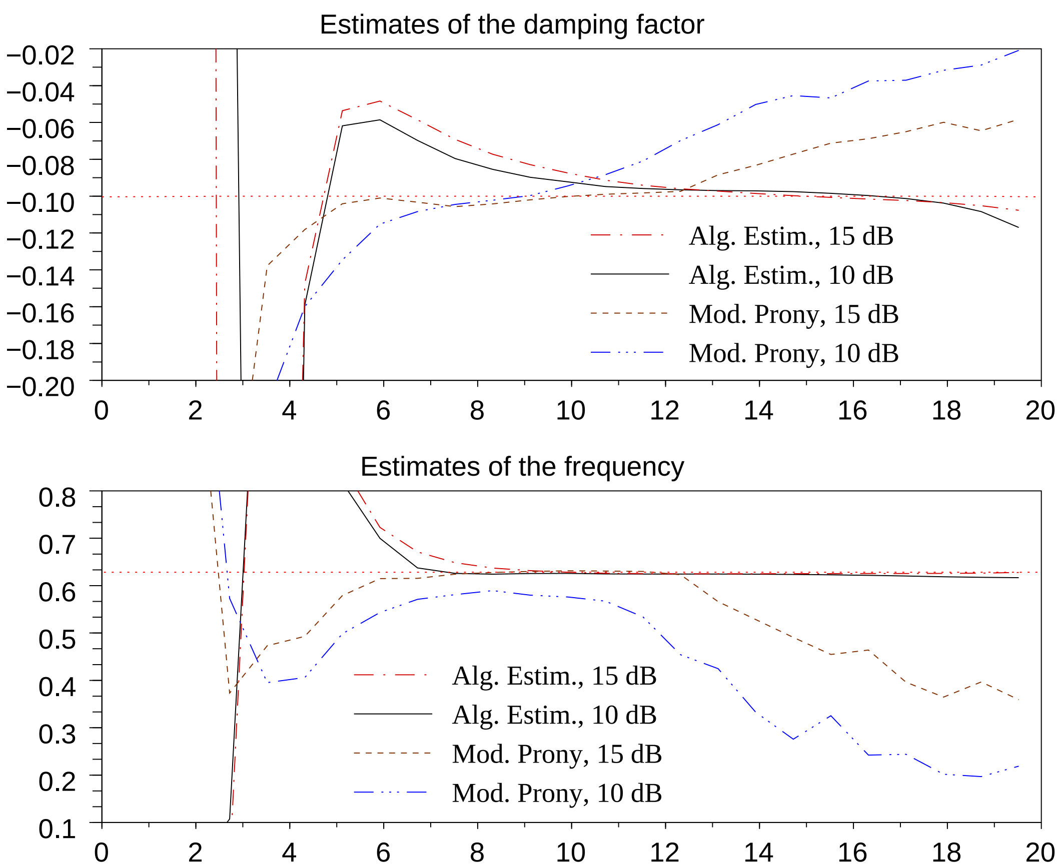


Fig. 3. Estimates of α and ω vs the estimation time.

Here again, the results obtained with the algebraic estimation, with $SNR = 10dB$ and $SNR = 15dB$ are very close to each other. And, of course, for the same $SNR = 15dB$, the performance are significantly better for $N = 256$ than for $N = 25$ (compare with figure 2). It appears, then, that the algebraic estimator is more robust to noise than the modified Prony's method. Indeed, if we increase further the number of points to have the numerical integration error dominated by the effect of the noise, then we still obtain little difference between the results with $SNR = 10dB$ and $SNR = 15dB$. The simulation results displayed in figure 4 are obtained with $N = 1024$.

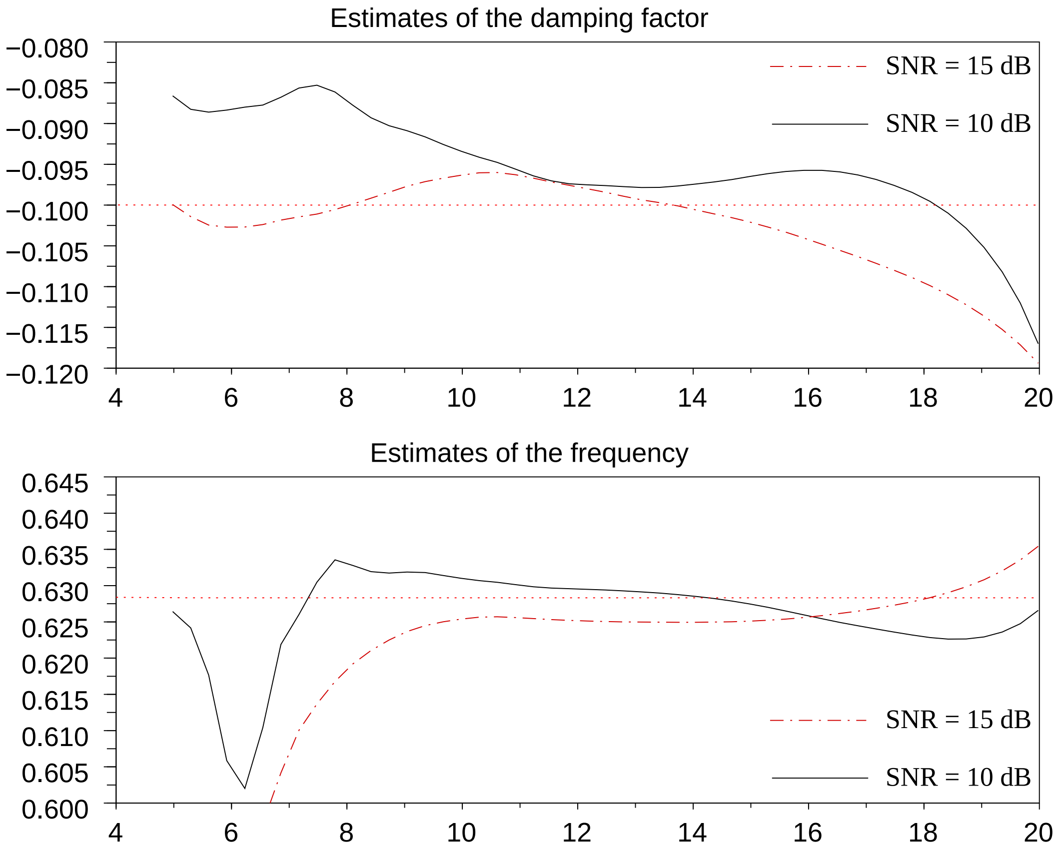


Fig. 4. Estimates of α and ω vs the estimation time.

Not only the estimates are more acurate than for $N = 256$ but also the performance are equivalent for both SNR. Since the performance of the modified Prony method degrade with small sampling periods, we did not simulate this method for $N = 250$ and $N = 1024$.

3 Mathematical background

The materials presented in this section are borrowed from [1] and [2]. They deal with differential algebra and operational calculus. Differential algebra, which is now widely used in control theory, provides us with powerful and elegant mean to exhibit simple hidden linear structures by allowing the coefficients to live in a much more rich ring/field. As a matter of fact, this idea is already well known in signal processing. Indeed, it is very classical to rewrite the differential equation

$$\sum_j a_j y^{(j)}(t) - \sum_i b_i x^{(i)} = 0,$$

representing the input $x(t)$ - output $y(t)$ relation of a linear time-invariant filter, as a linear combination

$$Ay(t) - Bx(t) = 0$$

in which the ring/field of scalars (coefficients) is now an extension of the usual field of real or complex numbers.

3.1 Differential algebra

To begin, we recall some basic definitions.

3.1.0.1 Rings and fields A *ring* is a set R with two law of compositions $(R, +, \cdot)$ such that

- $(R, +)$ is a commutative group,
- the multiplication is associative and distributive with respect to addition.

The ring R is called commutative if the multiplication law is commutative. A *field* is a commutative ring R in which every nonzero element is invertible in R.

3.1.0.2 Differential Ring/Field A *differential ring* R is a commutative ring which is equipped with a single derivation, written here $\frac{d}{ds}$, *i.e*, a map $R \to R$ such that, $\forall\ x, y \in R$,

- $\frac{d}{ds}(x+y) = \frac{dx}{ds} + \frac{dy}{ds}$,
- $\frac{d}{ds}(xy) = \frac{dx}{ds}y + x\frac{dy}{ds}$.

A *differential field* is a differential ring which is a field[4]. A *constant* $c \in R$ is such that $\frac{dc}{ds} = 0$. The set of all constants of a given differential ring (resp. field) is a differential subring (resp. subfield), called the *subring* (resp. *subfield*) *of constants*. A *(differential) ring* (resp. *field*) *of constants* is a differential ring (resp. field) whose elements are constant.

[4] All fields are assumed here to be of characteristic 0. See [26] for basic notions in commutative algebra.

Example 1 *The field* $\mathbb{Q}(s)$ *of rational functions in the indeterminate* s*, with coefficients in* $\mathbb{Q}$*, possesses an obvious structure of differential field with respect to* $\frac{d}{ds}$*. Its subfield of constants is* $\mathbb{Q}$*.*

3.1.0.3 Differential field extension

A *differential field extension* L/K is given by two differential fields K, L such that

- $K \subseteq L$,
- the restriction to K of the derivation of L is the derivation of K.

Example 2 $\mathbb{Q}$ *is a differential field with the derivation defined as the zero mapping. The restriction to* $\mathbb{Q}$ *of the derivation* $\frac{d}{ds}$ *of* $\mathbb{Q}(s)$ *is obviously the zero mapping:* $\mathbb{Q}(s)/\mathbb{Q}$ *is a differential field extension.*

An element $x \in L$ is said to be *differentially algebraic* over K if, and only if, x satisfies an algebraic differential equation over K, *i.e.*, $P(x, \frac{dx}{ds}, \ldots, \frac{d^n x}{ds^n}) = 0$, where P is a polynomial over K. The extension L/K is said to be *differentially algebraic* if, and only if, any element of L is differentially algebraic over K.

An element of L which is not differentially algebraic over K is said to be *differentially transcendental*. A *differentially transcendental* extension L/K is an extension which is not differentially algebraic.

3.1.0.4 Linear differential operators

Let K be a differential field. The set of linear differential operators of the form $\sum_{\text{finite}} a_i \frac{d^i}{ds^i}$, $a_i \in K$, is a differential ring, denoted by $K[\frac{d}{ds}]$. It is commutative if, and only if, K is a field of constants. To see this, take an element $a \in K$. Then, $\frac{d}{ds}a = \frac{da}{ds} + a\frac{d}{ds}$. Even in the general non-commutative case, it is known (see, *e.g.*, [27]) that $K[\frac{d}{ds}]$ is a principal left and right ideal ring: any left or right ideal may be generated by a single element.

3.2 The differential field of Mikusiński's operators

3.2.1 Mikusiński's field of operators

Endow the set $\mathcal{C}$ of continuous functions $[0, +\infty) \to \mathbb{C}$ with a structure of commutative ring with respect to the addition $(f + g)(t) = f(t) + g(t)$ and to the convolution (product) $(f \star g)(t) = (g \star f)(t) = \int_0^t f(\tau)g(t - \tau)d\tau = \int_0^t g(\tau)f(t-\tau)d\tau$. According to a famous theorem due to Titchmarsh (see [10–12,28]), $\mathcal{C}$ does not possess zero divisors. Any element of the *Mikusiński field* $\mathcal{M}$, *i.e.*, the quotient field of $\mathcal{C}$, is called an *operator*. Any function $f : \mathbb{R} \to \mathbb{C}$, which belongs to $\mathcal{M}$, may also be written $\{f\}$. Note that in general the product of two elements $a, b \in \mathcal{M}$ will be written ab and not $a \star b$. Some examples are in order:

(1) The neutral element $1 \in \mathcal{M}$ with respect to the convolution is the analogue of the Dirac measure at $t = 0$ in Schwartz's distribution theory.
(2) Any locally Lebesgue-integrable function $\mathbb{R} \to \mathbb{C}$ with a left bounded support belongs to $\mathcal{M}$.
(3) The inverse in $\mathcal{M}$ of the Heaviside function

$$\mathbf{1}(t) = \begin{cases} 0 & \text{if} \quad t < 0 \\ 1 & \text{if} \quad t \geq 0 \end{cases}$$

is the derivation operator s. Let $f : \mathbb{R} \to \mathbb{C}$ be a C^1-function with a left bounded support. Then $s\{f\} = \{f'\}$. Let $g : \mathbb{R} \to \mathbb{C}$ be a locally Lebesgue-integrable function with a left bounded support. Then $\frac{\{g\}}{s} = \{\int_{-\infty}^{t} g(\sigma) d\sigma\}$ has also a left bounded support. The meaning of the subfield $\mathbb{C}(s) \subset \mathcal{M}$ of rational functions over $\mathbb{C}$ in the indeterminate s is the usual one in operational calculus (see, *e.g.*, [10–12]).
(4) The meaning of the *delay* operator e^{-Ls}, $L \in \mathbb{R}$, is the usual one in operational calculus (see, *e.g.*, [10–12]). It is the analogue of the Dirac measure at $t = L$ in the theory of distributions.

3.2.2 The algebraic derivative

For any $f \in \mathcal{C}$, it is known (see [10–12]) that the mapping $f \mapsto \frac{df}{ds} = \{-tf\}$ satisfies the properties of a derivation, *i.e.*,

$$\frac{d}{ds}(f + g) = \frac{df}{ds} + \frac{dg}{ds}$$

and

$$\frac{d}{ds}(f \star g) = \frac{df}{ds} \star g + f \star \frac{dg}{ds}.$$

It can be trivially extended to a derivation, called the *algebraic derivative*, of $\mathcal{M}$ by setting, if $g \neq 0$,

$$\frac{d}{ds}(\{f\} \star \{g\}^{-1}) = \frac{\frac{df}{ds} \star g - f \star \frac{dg}{ds}}{\{g\}^2}$$

Endowed with the algebraic derivative, $\mathcal{M}$ becomes a differential field, whose subfield of constants is $\mathbb{C}$.

All fields are subfields of a differential field which is a *universal extension* [4] of the field $\mathbb{Q}$ of rational numbers.

3.3.1 The mathematical framework

Let k_0 be a given ground field, which is assumed to be a differential field of constants. Let k be a finite algebraic extension of $k_0(\boldsymbol{\Theta})$ where $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_r)$ is a finite set of *unknown parameters*. For example,

$$\lambda_0 + \frac{\lambda_1 \theta_1^2 \theta_2}{\lambda_2 + \lambda_3 \theta_2^3},$$

where the $\lambda_i' s, i = 0, \ldots, 3$ belong to k_0, is an element of k. Thus the transcendence degree of the extension k/k_0 does not exceed r. Moreover we give to k a canonical structure of a differential field of constants. Let $\mathcal{S}/k(s)$ be a finitely generated differentially algebraic extension. A *signal*, with parameter $\boldsymbol{\Theta}$, is an element of $\mathcal{S}$.

Consider now a finite set $\boldsymbol{x} = (x_1, \ldots, x_\kappa)$ of signals.

Definition 1 (Algebraic/rational) *The parameters* $\boldsymbol{\Theta}$ *are said to be* algebraically *(resp.* rationally*)* identifiable[5] *with respect to* $\boldsymbol{x}$ *if, and only if,* $\theta_1, \ldots, \theta_r$ *are algebraic over (belong to)* $k_0\langle s, \boldsymbol{x}\rangle$*. Here,* $N = \{s, \boldsymbol{x}\}$ *is a subset of* $\mathcal{S}$ *and* $k_0\langle s, \boldsymbol{x}\rangle$ *stands for the differential overfield of* k_0 *generated by* N*.*

Definition 2 (Linear identifiability) *The parameters* $\boldsymbol{\Theta}$ *are said to be* linearly identifiable *with respect to* $\boldsymbol{x}$ *if, and only if,*

$$P \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} = Q \tag{23}$$

where

- P *and* Q *are respectively* $r \times r$ *and* $r \times 1$ *matrices,*
- *the entries of* P *and* Q *belong to* $\mathrm{span}_{k_0(s)[\frac{d}{ds}]}(1, \boldsymbol{x})$,
- $\det(P) \neq 0$.

Here, $\mathrm{span}_{k_0(s)[\frac{d}{ds}]}(1, \boldsymbol{x})$ denotes the set of all linear combinations of the elements $1, x_1, \cdots, x_\kappa$ *i.e.* $D_0 \cdot 1 + \sum_{i=1}^{\kappa} D_i x_i$, where the coefficients D_i belong to

[5] Those definitions are borrowed from [29].

$k_0(s)[\frac{d}{ds}]$. Each D_i is thus a differential operator of the form

$$\sum_{j=0}^{\iota} \left(\frac{\sum_{n=0}^{n_j} a_{n,j} s^n}{\sum_{m=0}^{m_j} b_{m,j} s^m} \right) \frac{d^j}{ds^j}, \quad \text{with } a_{n,j}, b_{m,j} \in k_0.$$

Definition 3 (Projective/Weak linear) *The parameters* $\mathbf{\Theta}$ *are said to be*

- projectively linearly identifiable *with respect to* $\boldsymbol{x}$ *if, and only if, there exists* $\theta_\epsilon \neq 0$ *such that* $\frac{\theta_1}{\theta_\epsilon}, \ldots, \frac{\theta_{\epsilon-1}}{\theta_\epsilon}, \frac{\theta_{\epsilon+1}}{\theta_\epsilon}, \ldots, \frac{\theta_r}{\theta_\epsilon}$ *are linearly identifiable w.r.t.* $\boldsymbol{x}$.
- weakly linearly identifiable *with respect to* $\boldsymbol{x}$ *if, and only if, there exists a finite set* $\mathbf{\Theta}' = (\theta'_1, \ldots, \theta'_{q'})$ *such that*
  · *the components of* $\mathbf{\Theta}'$ *(resp.* $\mathbf{\Theta}$*) are algebraic over* $k_0(\mathbf{\Theta})$ *(resp.* $k_0(\mathbf{\Theta}')$*),*
  · $\mathbf{\Theta}'$ *is linearly identifiable.*

The next result is clear:

Proposition 1 *We have the following implications between the different kinds of identifiability*

$$\begin{array}{ccc} \textit{Linear} & \Longrightarrow & \textit{Rational} \\ \Downarrow & & \Downarrow \\ \textit{Weak linear} & \Longrightarrow & \textit{Algebraic} \end{array}$$

3.3.2 Rational signals

A *rational signal*

$$x = \frac{b_0 + b_1 s + \cdots + b_{n-1} s^{n-1}}{a_0 + a_1 s + \cdots + a_{n-1} s^{n-1} + s^n} = \frac{B(s)}{A(s)} \tag{24}$$

is an element of $k(s)$, where $k_0 = \mathbb{Q}$ and $k = k_0(a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1})$.

Proposition 2 *Assume that the numerator and the denominator of* (24) *are coprime. Then, the coefficients* $a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}$ *are linearly identifiable with respect to* x.

PROOF. Equation (24) yields the linear system of equations of type (23):

$$\frac{d^k}{ds^k} \left[B(s) - A(s)x \right] = \frac{d^k(s^n x)}{ds^k} \qquad k = 0, \ldots, 2n-1 \tag{25}$$

3.3.3 Differentially rational signals

A signal x is said to be *differentially rational* if, and only if, it satisfies $Dx = p$, where $D \in k(s)[\frac{d}{ds}]$ and $p \in k(s)$. It is easy to see that such a signal satisfies a differential equation of the form

$$\left(\sum_{\text{finite}} a_{\alpha\beta} s^{\alpha} \frac{d^{\beta}}{ds^{\beta}}\right) x = \sum_{\text{finite}} b_{\gamma} s^{\gamma}. \tag{26}$$

Setting $k_0 = \mathbb{Q}$, $k = k_0(a_{\alpha\beta}, b_{\gamma})$, the next result, which is a direct generalisation of proposition 2, may be proved in the same way.

Proposition 3 *Assume that in Equation 26 the polynomials $\sum_{\alpha} a_{\alpha\beta} s^{\alpha}$ and $\sum_{\gamma} b_{\gamma} s^{\gamma}$ are coprime. Then, the coefficients $a_{\alpha\beta}$, b_{γ} are projectively linearly identifiable with respect to x.*

3.3.4 Introducing exponentials

Lemma 1 *The expression*

$$x = \sum_{i=0}^{M} a_i \left(\frac{1}{s^{i+1}} - \sum_{j=0}^{i} \frac{\tau^j}{j!} \frac{1}{s^{i-j+1}}\right) + e^{-\tau s} \sum_{i=0}^{M} \frac{b_i}{s^{i+1}}$$

where

- $k_0 = \mathbb{Q}$, $k = k_0(a_i, b_i, \tau)$,
- $e^{-\tau s}$ *satisfies the differential equation* $(\frac{d}{ds} - \tau)e^{-\tau s} = 0$,

is differentially algebraic over $k(s)$.

PROOF. It follows from the fact that finite sums and products of differentially algebraic elements over $k(s)$ are again differentially algebraic over $k(s)$.

Proposition 4 *The parameters τ, a_i, b_i, $i = 0, \ldots, M$, are algebraically identifiable with respect to the signal x.*

PROOF. Take as in the proof of theorem 2 sufficiently many derivatives of $s^M x$ with respect to s. The conclusion follows from the transcendence [6] of $e^{-\tau s}$ over $k(s)$.

[6] See [30] for a direct proof without having recourse to analytic functions.

3.4.1 Structured perturbations

Let k_1/k_0 be a differential field extension such that

- k_1 is a differential field of constants,
- k and k_1 are linearly disjoint over k_0.

Definition 4 *A* perturbation ϖ *is an element of a differential overfield* $\mathcal{N}$ *of* $k_1(s)$ *such that* $\mathcal{S}$ *and* $\mathcal{N}$ *are linearly disjoint over* $k_0(s)$*. It is said to be* structured *if, and only if, it is annihilated by* $\Pi \in k_0(s)[\frac{d}{ds}]$, $\Pi \neq 0$.

Example 3 *Consider the perturbation* $\frac{\gamma}{s^\nu}$, $\gamma \in k_1$*. It is annihilated by* $\nu s^{\nu-1} + s^\nu \frac{d}{ds} \in k_0(s)[\frac{d}{ds}]$*, which does not depend on* γ*.*

3.4.2 Noise perturbation

A perturbation which is not structured is said to be *unstructured*. An unstructured perturbation is called a *noise*. In practice we will assume that when specialized to $\mathcal{M}$ a noise corresponds to a rapidly oscillating time-function, *i.e.* a “high frequency” signal, which may be attenuated by a low pass filter. A detailed analysis of the noise effect is presented in [31], using the formalism of nonstandard analysis.

3.4.3 Noisy signals

A *signal with an additive noise* is a sum $x + \varpi$, where $x \in \mathcal{S}$ is a signal and $\varpi \in \mathcal{N}$ a noise. Let $\boldsymbol{y} = (y_1, \ldots, y_\kappa)$, where $y_\iota = x_\iota + \varpi_\iota$ be a finite set of such noisy signals. If the parameters $\boldsymbol{\Theta}$ are linearly identifiable, then equation (23) becomes

$$P \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} = Q + Q' \tag{27}$$

where

- the matrices P and Q are obtained from (23) by substituting $\boldsymbol{y}$ to $\boldsymbol{x}$,
- the entries of the $r \times 1$ vector Q' belong to $\mathrm{span}_{k'(s)[\frac{d}{ds}]}(\boldsymbol{\varpi})$, where k' is the quotient field of $k \otimes_{k_0} k_1$, and $\boldsymbol{\varpi} = (\varpi_1, \ldots, \varpi_\kappa)$.

3.4.4 Linear estimators

Assume that the components of $\boldsymbol{\varpi}$ are structured. The next fundamental theorem follows at once from the fact that $k_0(s)[\frac{d}{ds}]$ is a principal left ideal ring.

Theorem 1 *There exists* $\Pi \in k_0(s)[\frac{d}{ds}]$, *such that equation (27) becomes*

$$\Pi P \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} = \Pi Q \tag{28}$$

Equation (28), which is independent of the noises, is called a *linear estimator* of the unknown parameters if, and only if, $\det(\Pi P) \neq 0$. On the other hand, if the components of $\boldsymbol{\varpi}$ are not structured, then the right hand side of (28) would read as $\Pi Q + \Pi Q'$. Ignoring the additional term $\Pi Q'$, results in an approximation of $\boldsymbol{\Theta}$ by $\widetilde{\boldsymbol{\Theta}}$, the solution of the system $\Pi P \widetilde{\boldsymbol{\Theta}} = \Pi Q$. In the sequel, we will absorbe Π and write P and Q in place of ΠP and ΠQ, respectively.

Let ℓ be a differential field of constants. A differential operator in $\ell(s)[\frac{d}{ds}]$ is said to be *proper* (resp. *strictly proper*) if, and only if, the coefficients of $\frac{d^\alpha}{ds^\alpha}$ are proper (resp. strictly proper) rational functions in $\ell(s)$. The estimator (28) is said to be *proper* (resp. *strictly proper*) if, and only if, the entries of ΠP and ΠQ are proper (resp. strictly proper) differential operators. Multiplying both sides of equation (28) by a suitable proper element of $k_0(s)$ yields the

Proposition 5 *Any linear estimator may be replaced by a proper (resp. strictly proper) one.*

4 Analysis

Some basic features of the presented estimation method are discussed in this section.

4.1 Least squares interpretation

As a matter of fact, the principle of the estimation algorithm presented so far is connected with that of orthogonal projection. An interpretation in terms of least squares follows then.

To see this, let us revisit the main steps leading to a linear estimator of $\mathbf{\Theta}$, from the observation $y = F(x, \mathbf{\Theta}) + n(t)$, as presented in section 2.

The first step was to find a differential operator $D_{\mathbf{\Theta}}$ such that the unobserved signal x satisties $D_{\mathbf{\Theta}} x = p$, where p represents the contribution of the structured perturbations. This differential equation induces an algebraic map $\xi : \mathbb{C}^r \to \mathbb{C}^q$,

$$\mathbf{\Theta} = (\theta_1, \ldots, \theta_r) \mapsto \mathbf{\Theta'} = (\theta'_1, \ldots, \theta'_q) = \xi(\mathbf{\Theta}), \tag{29}$$

which reduces to an identity in case of linear identifiability. Translated into the operational domain, that is in $\mathcal{M}$, one gets an equation of the form

$$\sum_{i=1}^{q} \theta'_i \hat{f}_i(s) = \hat{g}(s) + \hat{p}(s). \tag{30}$$

Here, $\hat{g}(s)$ and $\hat{f}_i(s), i = 1, \ldots, \kappa$, are of the form[7] $d + D\hat{x}$, with $d \in k_0(s)$ and $D \in k_0(s)[\frac{d}{ds}]$ and $\hat{p}(s) \in k(s)$. We assume, without any loss of generality, that $\hat{p}(s)$ is a polynomial over k. This can always be achieved by simply multiplying both sides of (30) by the denominator of $\hat{p}(s)$. We now replace x by the observation y and redefine $\hat{g}(s)$ and $\hat{f}_i(s), i = 1, \ldots, \kappa$, according to this modification. It is clear that then, the equality in (30) does no longer hold. So, let us introduce the "error" function

$$\hat{e}(s\,; \boldsymbol{\vartheta}, \hat{y}) = \sum_{i=1}^{q} \vartheta'_i \hat{f}_i(s) - \hat{g}(s) - \hat{p}(s), \tag{31}$$

parametrized by $\hat{y}$ (to quote the substitution of $\hat{x}$ by $\hat{y}$) and by $\boldsymbol{\vartheta'} = \xi(\boldsymbol{\vartheta})$ where $\boldsymbol{\vartheta}$ is a vector of free parameters. In particular, equation (30) corresponds to

$$\hat{e}(s\,; \mathbf{\Theta}, \hat{x}) = 0.$$

At this stage, it is not difficult to see that steps 1-3 of the algorithm presented in section 2.1.2 may be restated as follows: a (strictly) proper linear estimator, $\widetilde{\mathbf{\Theta}}' = \xi(\widetilde{\mathbf{\Theta}})$, of $\mathbf{\Theta'} = \xi(\mathbf{\Theta})$ is given by the system of equations:

$$\frac{1}{s^\nu} \frac{d^m}{ds^m} \hat{e}(s\,; \widetilde{\mathbf{\Theta}}, \hat{y}) = 0\ ;\ m = \kappa, \ldots, q + \kappa - 1, \tag{32}$$

where $\kappa \geq \deg \hat{p}(s) + 1$ is the order of the additional differentiations for eliminating the contribution of the structured perturbations. In the sequal, such a proper linear estimator will be called an *algebraic estimator*. Setting $e(t\,; \widetilde{\mathbf{\Theta}}, y)$ for the time domain analogue of $\hat{e}(s\,; \widetilde{\mathbf{\Theta}}, \hat{y})$, this system (32) reads back in the

[7] In general either d of D is zero.

time domain, as

$$
\begin{aligned}
0 &= \frac{(-1)^m}{(\nu-1)!}\int_0^T (T-\tau)^{\nu-1}\tau^m e(\tau\,;\widetilde{\boldsymbol{\Theta}}, y(\tau))d\tau \\
&= \frac{(-1)^m}{(\nu-1)!}\int_0^T \{(T-\tau)^{\nu-1}\tau^\kappa\}\tau^{m-\kappa} e(\tau\,;\widetilde{\boldsymbol{\Theta}}, y(\tau))d\tau \\
m &= \kappa, \ldots, q+\kappa-1,
\end{aligned}
\tag{33}
$$

where T denotes the estimation time. Let us focus on the bracketed term under the integral in the second line of (33). It is of the form

$$
w_{(\nu,\kappa)}(t) \triangleq (T-t)^{\alpha-\beta} t^{\beta-1}, \tag{34}
$$

with $\beta = \kappa + 1 > 0$ and $\alpha = \nu + \kappa$ and interestingly, one recognizes the weight function associated to the orthogonal Jacobi polynomials $\{P_j^{(\nu,\kappa)}(t)\}_{j\geqslant 0}$, in $[0, T]$. This observation will play an important role in what follows. To proceed, note that we may readily rewrite the system of equations (33) as

$$
0 = \left\langle t^j,\, e(t\,;\widetilde{\boldsymbol{\Theta}}, y)\right\rangle_{w_{(\nu,\kappa)}}, \tag{35}
$$

$$
\Longleftrightarrow\ 0 = \left\langle P_j^{(\nu,\kappa)}(t),\, e(t\,;\widetilde{\boldsymbol{\Theta}}, y)\right\rangle_{w_{(\nu,\kappa)}},\ j = 0, \ldots, q-1, \tag{36}
$$

where, for $f(t)$ and $g(t)$ two real fonctions defined in $[0, T]$, their scalar product with respect to $w_{(\nu,\kappa)}(t) \geq 0$, $\forall t \in [0, T]$ is given by:

$$
\langle f, g\rangle_{w_{(\nu,\kappa)}} \triangleq \int_0^T f(t)g(t)w_{(\nu,\kappa)}(t)dt.
$$

The following theorem, which summarizes the preceding developpements, is therefore proved.

Theorem 2 *Let $y(t) = F(x, \boldsymbol{\Theta}) + n(t)$ denotes the noisy observation, through a functional F, of the signal x which depends on the parameters $\boldsymbol{\Theta}$. Assume that the parameters $\boldsymbol{\Theta}$ are weakly linearly identifiable with respect to x, through the algebraic map* (29)*. Then, $\widetilde{\boldsymbol{\Theta}}$ is an algebraic estimate of $\boldsymbol{\Theta}$ in $[0, T]$ if, and only if, it satisfies the orthogonality condition*

$$
e(t\,;\widetilde{\boldsymbol{\Theta}}, y) \in \left[span\left\{P_0^{(\nu,\kappa)}(t), \ldots, P_{q-1}^{(\nu,\kappa)}(t)\right\}\right]^{\perp}. \tag{37}
$$

4.2 Estimation time

As already quoted, the estimation time T may be small especially in the absence of noise, *i.e.* unstructured perturbation. Meanwhile, T can not, obviously, be taken arbitrary small even in a noise-free context. In this connexion, a lower bound for T has been formally characterized in [31, Prop. 3.2], within the framework of nonstandard analysis. Here, we give another and hopefully more explicit description of this lower bound. We begin with equation (30) where (without any loss of generality) we ignore $\hat{p}$, the contribution of the structured perturbations. So, let $f_j(t), j = 1, \ldots, q$ and $g(t)$, denotes respectively, the time domain analogues of $\hat{f}_j(s), j = 1, \ldots, q$ and $\hat{g}(s)$ therein. Assume that these are continuous functions in $[0, T]$. Then due to the Weierstrass approximation theorem, we may write, for $t \in [0, T]$ and for any $\varepsilon > 0$,

$$f_j(t) = \sum_{i=0}^{N_j(\varepsilon,T)} f_{i,j} P_i^{(\nu,\kappa)}(t) + \varepsilon_j(t) \tag{38a}$$

$$g(t) = \sum_{i=0}^{N_0(\varepsilon,T)} g_i P_i^{(\nu,\kappa)}(t) + \varepsilon_0(t) \tag{38b}$$

where $|\varepsilon_j(t)| < \varepsilon$, $\forall t \in [0, T]$, $j = 0, \ldots, q$. In the following, we consider that ε is sufficiently small so that it can be neglected. Accordingly, we will set $N_j(T)$ for $N_j(\varepsilon, T)$ to ease the notations. Equation (30) is then equivalent (up to ε), in the time domain for $t \in [0, T]$, to:

$$\sum_{j=1}^{q} \left(\sum_{i=0}^{N_j(T)} f_{i,j} P_i^{(\nu,\kappa)}(t) \right) \theta'_j = \sum_{i=0}^{N_0(T)} g_i P_i^{(\nu,\kappa)}(t). \tag{39}$$

In matrix form, this equation becomes

$$\begin{bmatrix} & \mathcal{F}_q & \\ & & \\ \cdots & \cdots & \cdots \\ f_{q,1} & \cdots & f_{q,q} \\ \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} \theta'_1 \\ \vdots \\ \theta'_q \end{bmatrix} = \begin{bmatrix} \mathcal{G}_q \\ \\ \cdots \\ g_q \\ \vdots \end{bmatrix} \tag{40}$$

Replace x by y and accordingly, we will write $\mathcal{F}_q(T, y)$ and $\mathcal{G}_q(T, y)$ in order to make explicit the dependance of the matrices in (40) on T and y. Then the following proposition is plain.

Proposition 6 *Assume that the parameters* $\boldsymbol{\Theta}$ *are weakly linearly identifiable with respect to* x*, through the algebraic map* (29)*. Then,* $\widetilde{\boldsymbol{\Theta}}$ *is an algebraic estimate of* $\boldsymbol{\Theta}$ *in* $[0, T]$ *if, and only if,*

- *the estimation time* T *is such that* $\det \mathcal{F}_q(T, y) \neq 0$,
- $\mathcal{F}_q(T, y)\widetilde{\boldsymbol{\Theta}'} = \mathcal{G}_q(T, y)$.

Corollary 1 *A lower bound for the estimation time T is given by*

$$T_{min} = \arg\min_{T} \{N_{max}(T) \geq q-1\}\,. \tag{41}$$

where $N_{max}(T) = \max_{1\leqslant j\leqslant q} N_j(T)$.

Indeed, if T is smaller that the T_{min} above then the last row of $\mathcal{F}_q(T,y)$ is zero.

Remark 2 *This corrolary may be interpreted in terms of "sufficient excitation condition", a generic condition in any estimation problem.*

Of course, in a noisy setting, an additional estimation time is required to filter out the noise. But this estimation time can not be too large either. To see this, recall, from equations (31)-(33), that a linear numerical estimator of $\mathbf{\Theta}$ is of the form $\mathcal{P}_\nu(T)\widetilde{\mathbf{\Theta}} = \mathcal{Q}_\nu(T)$ with

$$\{\mathcal{P}_\nu(T)\}_{i,j} = \int_0^T (T-\tau)^{\nu-1}\tau^{\kappa-1+i} f_j(\tau)d\tau,$$

and likewize for $\{\mathcal{Q}_\nu(T)\}_i$ with f_j replaced by g. Each of these entries then appears as the output at time T, of the linear time-invariant causal and *unstable* filter with impulse response $h_\nu(t) = t^{\nu-1}, \nu \geqslant 1$, for the input signals $t^{\kappa-1+i}u(t)$, with $u(t) = f_j(t)$ and $u(t) = g(t)$ respectively.

Before closing this subsection, the following remarks are in order:

- Any argument of asymptotic type ($T \to \infty$) is to be banished in the performance analysis of the estimators designed here within the algebraic framework.
- There exists an optimal value for the estimation time, corresponding to a minimum mean square estimation error.

4.3 Noise effect

Assume that the parameters $\mathbf{\Theta}$ are linearly identifiable with respect to x *i.e.* $P\mathbf{\Theta} = Q$. Replacing x by $y = x + n$, when an unstructured perturbation n is present, yields the estimates $\widetilde{\mathbf{\Theta}}$ as the solution [8] of

$$(P + \Delta P)\widetilde{\mathbf{\Theta}} = (Q + \Delta Q). \tag{42}$$

[8] Compare with (27) which describes the true parameter vector $\mathbf{\Theta}$. Indeed, since the term Q' in the second member of (27) depends on $\mathbf{\Theta}$, that equation is not implementable as a linear estimator.

From numerical linear algebra, we have the classical first order bound [32]

$$\frac{\|\mathbf{\Theta} - \widetilde{\mathbf{\Theta}}\|}{\|\mathbf{\Theta}\|} \leqslant \frac{2\varepsilon K(P)}{1 - \varepsilon K(P)}, \tag{43}$$

where ε measures the level of the noise and the matrix condition $K(P) = \|P^{-1}\| \|P\|$ determines the sensitivity of the system (23).

4.3.1 Output noise level

Among the various way of measuring the output noise level, we consider here the normwise metric [32]

$$\varepsilon = \max\left\{\frac{\|\Delta P\|}{\|P\|}, \frac{\|\Delta Q\|}{\|Q\|}\right\}.$$

with L_2-vector norm and the induced spectral norm for matrices. This noise measure is thus related to the noise-to-signal ratio.

Let us assume first that the additive noise $n(t)$ is a rapidly oscillating function with zero mean. Then, the analysis in [31] shows that the iterated integrals

$$\int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_1} \tau^\alpha n(\tau) d\tau dt_1 \cdots dt_{k-1}$$

are small even for small values of t. Since the entries of ΔP and ΔQ are linear combinations of terms of this form, we deduce that ε will be neglectable in this case. Unfortunately, the noise characteristics do not always fit into the context of [31]. This happens *e.g.* with the last scenario discussed in paragraph 2.2.1, where the noise do not corrupt the signal in a simple additive way. Anyway, the iterated integrals provide a low pass filtering which attenuates the noise effect. Starting to increase the order ν of the iterated integrals, from its minimum value guaranteing strict properness, amounts to reducing the corresponding low pass cutoff frequency. This reduces the output noise level. However, if we iterate too much the integrals, the output noise may increase, as compared with the signal. Indeed, if the cutoff frequency is too low, then the signal also may be affected. Another way to see this is to rewrite the iterated integral above, using the Cauchy formula for repeated integration (10). Doing so, makes intervene a weighting function $w_{(\nu,\kappa)}(t)$ of the form given in (34). Clearly, the effective estimation time length is determined by the significant part of the support of the weighting function $w_{(\nu,\kappa)}(t)$. Now, increasing ν reduces this length. Unless the noise is rapidly oscillating, its ouput, which results from the averaging afforded by the integration, may be significant as compared to the signal counterpart for small (effective) integration time.

To illustrate these facts, let us reconsider the system (19) in the estimation problem of section 2 and the second scenario therein (see paragraph 2.2.1) with an input SNR of $10dB$. We fix the estimation time to $t = T$, corresponding to the first 180 samples of the signal and we let the order ν of the iterated integrals increase from $\nu_0 = 4$, the minimum value guaranteing strict properness. Figure 5 below represents the averages over 100 trials of the output noise-to-signal ratios $\frac{\|\Delta\mathcal{P}_\nu(T)\|}{\|\mathcal{P}_\nu(T)\|}$ (solid line) and $\frac{\|\Delta\mathcal{Q}_\nu(T)\|}{\|\mathcal{Q}_\nu(T)\|}$ (dashed line), respectively, as a function of ν.

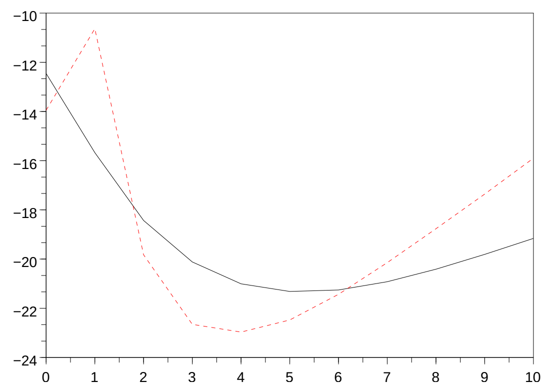


Fig. 5. Output noise level vs $\nu - \nu_0$, in dB

These results confirm the necessity of achieving a compromise in the selection of ν, especially as the optimal (in the minimum output noise power sense) values of ν for $\mathcal{P}_\nu(T)$ and for $\mathcal{Q}_\nu(T)$ do not coincide in general.

4.3.2 Sensitivity of the system

The second important factor in the noise effect analysis is, of course, the sensitivity of the system which tells how the noise affects the solution. Associated with the normwise metric used above, is the relative normwise condition number K_N defined, for a square system $A\boldsymbol{x} = \boldsymbol{b}$, by

$$K_N(A) = K(A) + \frac{\|A^{-1}\|\|\boldsymbol{b}\|}{\|\boldsymbol{x}\|} \tag{44}$$

where $K(\cdot)$ is the familiar matrix condition number as above. In the next figure, we have plotted the values of $K_N(\mathcal{P}_\nu(T))$ computed from the preceding simulation, againts ν.

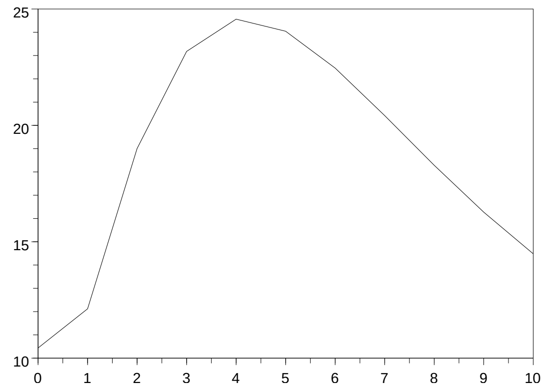


Fig. 6. Relative normwise condition number $K_N(\mathcal{P}_\nu(T))$ vs $\nu - \nu_0$.

The variations of the system's sensitivity are exactly in opposition to that of the output noise level. And determining the optimum value for ν (in the sense of achieving the best compromise), is still an open question.

Nonetheless, we mention below a simple trick to reduce this sensitivity. Let us begin with the

Lemma 2 *Assume that* $\mathbf{\Theta}$ *is (weakly) linearly identifiable with respect to* x. *Let the estimates* $\widetilde{\mathbf{\Theta}}$ *of* $\mathbf{\Theta}$ *be obtained, for an estimation time* T, *by the linear system*

$$\mathcal{P}_\nu(T)\widetilde{\mathbf{\Theta}} = \mathcal{Q}_\nu(T).$$

If $T < 1$, *then the system matrix* $\mathcal{P}_\nu(T)$ *is badly scaled.*

PROOF. To see this, let us recall the most general form of our estimators, according to the least squares interpretation discussed in section 4.1. By a change of variable in (33), one may readily rewrite the ℓ^{th} row ($\ell = 0, \ldots, q-1$) of the above system in the equivalent form

$$T^{\nu+\ell} \sum_{i=1}^{q} \left\{ \int_0^1 (1-\tau)^{\nu-1} \tau^\ell f_i(T\tau) d\tau \right\} \theta_i' = T^{\nu+\ell} \int_0^1 (1-\tau)^{\nu-1} \tau^\ell g(T\tau) d\tau. \quad (45)$$

Therefore, it is easy to show that the matrix $\mathcal{P}_\nu(T)$ and the vector $\mathcal{Q}_\nu(T)$ can be factored as: $\mathcal{P}_\nu(T) = \Lambda(T)\widetilde{\mathcal{P}}_\nu(1)$ and $\mathcal{Q}_\nu(T) = \Lambda(T)\widetilde{\mathcal{Q}}_\nu(1)$, where $\Lambda(T)$ is the diagonal matrix with $\{\Lambda(T)\}_{i,i} = T^{\nu+i-1}$, $i = 1, \ldots, q$.

Since the relative normwise condition number is scale dependent, a high system sensitivity may be unnaturally induced by the diagonal factor $\Lambda(T)$. Nonetheless, the reduced system $\widetilde{\mathcal{P}}_\nu(1)\widetilde{\mathbf{\Theta}} = \widetilde{\mathcal{Q}}_\nu(1)$ is still badly scaled: writing $\widetilde{\mathcal{P}}_\nu(1) = \int_0^1 \boldsymbol{p}_\nu(\tau) d\tau$, we have $|\{\boldsymbol{p}_\nu(\tau)\}_{i,j}| > |\{\boldsymbol{p}_\nu(\tau)\}_{\ell,j}|$, $\forall\, \ell > i$. The ill-conditioning is thus due to more than a simple scaling.

Recall that in the operational domain, our linear estimators have the general form

$$H(s)\widehat{P}(s)\widetilde{\boldsymbol{\Theta}} = H(s)\widehat{Q}(s) \tag{46}$$

where $H(s) \in k_0(s)$ is chosen so as to ensure strict properness. So far, we have set $H(s) = \frac{1}{s^\nu} I$. It turns out that the system's sensitivity may be significantly reduced by a judicious choice for $H(s)$.

In particular, a simple choice which is appropriate for balancing the rows of the system is given by

$$H(s) = \begin{bmatrix} \frac{1}{s^{\nu+q}} & & & \\ & \frac{1}{s^{\nu+q-1}} & & \\ & & \ddots & \\ & & & \frac{1}{s^\nu} \end{bmatrix} \tag{47}$$

Observe in figure 7, how K_N of the above experiment is improved by such a simple choice for $H(s)$. The corresponding relative condition number, associated with the componentwise model of Skeel [33] (see also [32] and [34]), is also displayed to avoid a possible scaling-induced miss-intrepretation. This condition number, defined for a linear square system $A\boldsymbol{x} = \boldsymbol{b}$, by

$$K_C(A) = \frac{\| |A^{-1}|(|A||\boldsymbol{x}| + |\boldsymbol{b}|)\|_\infty}{\|\boldsymbol{x}\|_\infty}, \tag{48}$$

is invariant to diagonal row scaling. Here the absolute value of the matrix $|A|$ is to be understood entrywise.

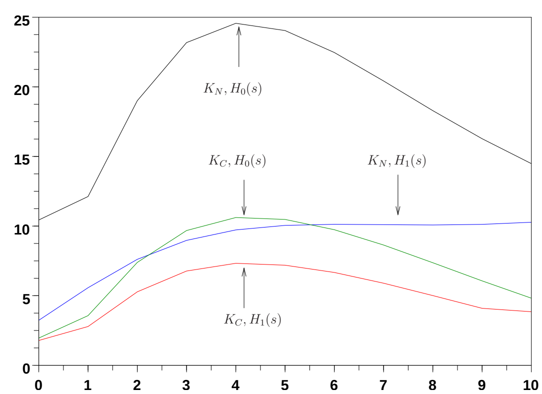


Fig. 7. Improved system's sensitivity: $H_0(s) = s^{-\nu} I$, $H_1(s) = \mathrm{diag}\,[s^{-(\nu+1)},\, s^{-\nu}]$.

Finally, note that the conclusion of Lemma 2 is also valid for $T > 1$. In that case, $H(s)$ should be chosen as $H(s) = \mathrm{diag}[s^{-\nu},\, s^{-(\nu+1)},\, \cdots,\, s^{-(\nu+q)}]$.

5 Applications

In this section, we illustrate further the behavior of our estimation method through two application examples.

5.1 *Chirp parameter estimation*

5.1.1 *Signal model and algorithm*

We still consider the estimation of the parameters of a complex exponential signal

$$x(t) = a \exp i\varphi(t), \tag{49}$$

where the phase function $\varphi(t)$ is now nonlinear. Such an estimation problem is important because signals of this type are encountered in many applications including radar, sonar, bioengineering, speech modeling, optics, gravity waves to name fews. It has a long and rich history that can be gauged by the great diversity of the available approaches (Maximum likelihood [35], subspace [36], wavelet and time-frequency [37] ...).

Our starting point is to observe that $x(t)$ fulfills the first order linear time-varying differential equation

$$\dot{x}(t) = i\dot{\varphi}(t)x(t). \tag{50}$$

Most often, the phase function is represented as

$$\varphi(t) = \varphi_0 + \varphi_1 t + \varphi_2 t^2,$$

resulting in a polynomial phase signal or linear chirp signal (see [2] for a treatment of the real case). With this representation, it appears clearly that $\hat{x}$, the analogue of $x(t)$ in $\mathcal{M}$, is the differentially rational signal described by

$$\left(-i\varphi_1 + s - 2i\varphi_2 \frac{d}{ds}\right)\hat{x} = x_0. \tag{51}$$

The parameter vector

$$\mathbf{\Theta} = (a,\, \varphi_0,\, \varphi_1,\, \varphi_2).$$

is thus weakly linearly identifiable with respect to x. This stems from the linear identifiability of

$$\mathbf{\Theta'} = (x_0,\, \varphi_1,\, \varphi_2)$$

by proposition 3. The estimates of $\mathbf{\Theta}$ are thus obtained by following the steps of the algorithm in section 2.1.2.

Note in passing that the hyperbolic chirp case, in which the phase function is of the form

$$\varphi(t) = \frac{\varphi_0}{\varphi_1 - t}, \quad 0 \leqslant t < \varphi_1,$$

may be handled in a very similar way. Indeed, in this case, (50) specialises to

$$(\varphi_1 - t)^2 \dot{x}(t) = i\varphi_0 x(t),$$

from which we deduce

$$\left([2\varphi_1 - i\varphi_0] + \varphi_1^2 s + 2\varphi_1 s \frac{d}{ds} - 2\frac{d}{ds} - s\frac{d^2}{ds^2} \right) \hat{x} = \varphi_1^2 x_0. \tag{52}$$

By proposition 3, $\mathbf{\Theta'} = (\varphi_1^2 x_0,\, [2\varphi_1 - i\varphi_0],\, \varphi_1,\, \varphi_1^2)$ is linearly identifiable with respect to x, and hence $\mathbf{\Theta} = (a,\, \varphi_0,\, \varphi_1)$ is weakly linearly identifiable.

5.1.2 Numerical simulation

The next figure depicts a sample realization of the real part of the noisy linear chirp signal (49), with parameters

$$\mathbf{\Theta} = (a,\, \varphi_0,\, \varphi_1,\, \varphi_2) = (2.291,\, 1.1766,\, 1.37,\, -3.95).$$

The sampling period is $T_s = 0.01s$, corresponding to $N = 200$ simulated

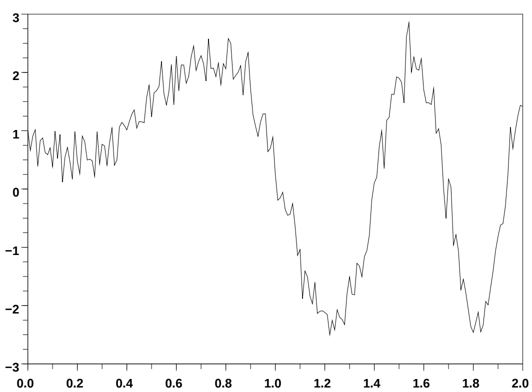


Fig. 8. Real part of the noisy observation signal - $SNR = 15dB$.

points. The additive noise is simulated as $n(t_k) = n_r(k) + in_i(k)$ where the real and imaginary parts are indepedent zero-mean white Gaussian noises with a variance corresponding to a SNR of $15dB$. Figure 9 below shows the estimates $\widetilde{\mathbf{\Theta}}(T)$, for 40 equally spaced values of T in $[0.4\,,2]$. The exact values of the parameters are also depicted with dashed lines, for comparaison. Each plot represents the results averaged over 1000 realizations. The estimates are unbiased for all T in the considered interval. Figure 10 displays the corresponding variances. Comparing with the associated Cramer-Rao lower bounds (dashed line curves) allows one to gauge the robustness of the estimator.

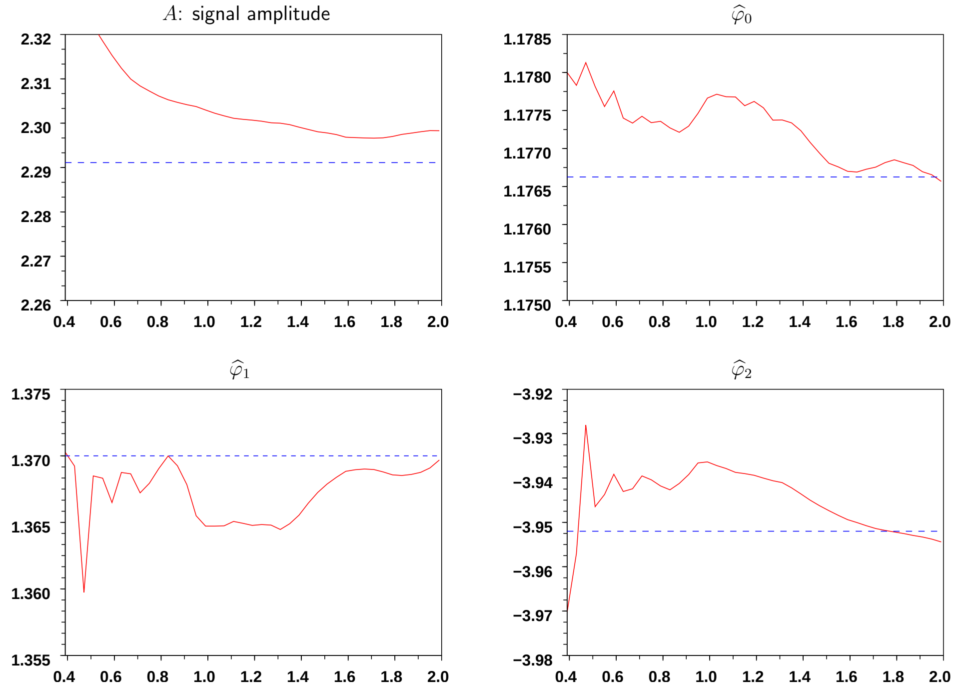


Fig. 9. Estimates of the chirp parameters *vs* estimation time - $SNR = 15dB$.

These curves also illustrate a basic feature of the presented algebraic estimators: There exists an optimal choice for the estimation time. Determing this optimum choice is however yet an open question.

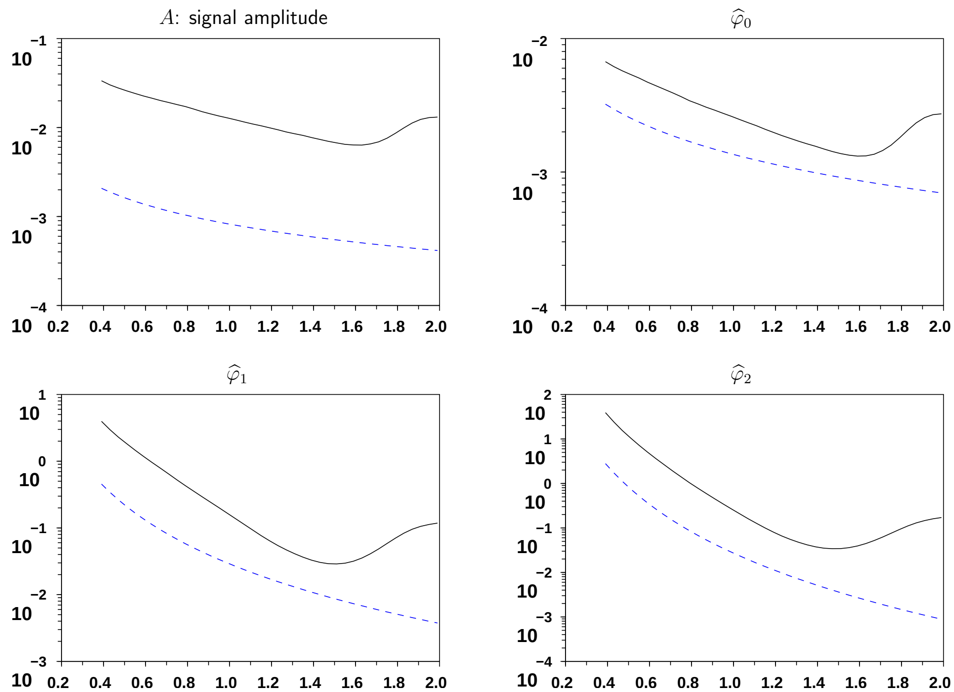


Fig. 10. Variance of the estimates of the chirp parameters vs estimation time - $SNR = 15dB$.

All the estimation problems considered so far rely on the knowledge of a precise parametric model. In this section, we show how the developped methodology may be accomodated to a situation where such a precise parametric model is not available.

Given a piecewise regular signal, the problem under concern is to estimate the locations of its discontinuities. This problem, known as the change point detection problem, is widely studied in the litterature. The Bayesian framework and the wavelet based approaches play a prominent part in the existing methods. These may be classified either as sequential [13], local [38] or global [39].

The solution we are presenting here is based on the estimation of the point of singularity of the signal's derivative. It is a local method and it is close in spirit to the wavelet based approaches [40], [41]. Following this spirit, one immediate solution is to use a signal derivative estimator, as those proposed in [42], to locate the singularities. But, as we will shortly see, an explicit estimation of the derivative is not necessary: the problem is rather casted into a delay estimation problem.

To begin, consider the following signal model,

$$y(t) = \sum_{i=1}^{K} \mathbb{1}_{[t_{i-1},t_i]}(t)\, p_i(t - t_{i-1}) + n(t), \tag{53}$$

$n(t)$ is an additive noise corruption. We set $t_0 = 0$ and $t_i, i = 1, \ldots, K$ are the change points. We denote by $x(t) = y(t) - n(t)$ the unobserved noise-free signal. Let us assume, for a moment, that each $p_i(t)$ is a polynomial and let T be given such that there is at most one discontinuity point in each interval $I_\tau^T = (\tau, \tau + T)$, $\tau \geqslant 0$. In the sequel, we will set

$$x_\tau(t) = x(t + \tau), \quad t \in [0, T],$$

for the restriction of the signal in I_τ^T and redefine the discontinuity point, say t_τ, relatively to I_τ^T with: $t_\tau = 0$ if $x_\tau(t)$ is smooth and $0 < t_\tau \leqslant T$ otherwise. Now, we know that the N^{th} order derivative of $x_\tau(t)$ (in the sense of distributions theory [43]) satisfies

$$\frac{d^N}{dt^N} x_\tau(t) = [x_\tau^{(N)}](t) + \sum_{k=1}^{N} \mu_{N-k}\, \delta(t - t_\tau)^{(k-1)} \tag{54}$$

where the superscript (k) denotes the derivation of order k. Here, μ_k is the jump on the k^{th} order derivative at the point t_τ and $[x_\tau^{(N)}]$ represents the regular part of the N^{th} order derivative of the signal. Several change point estimators may be devised from this equation, as shown below.

5.2.1 Estimator A

First, consider the simple case $N = 1$ in equation (54) and assume that the regular part $[\dot{x}_\tau](t)$ may be represented in each interval I_τ^T by a polynomial of degree not exceeding, say, $\kappa - 1$. This corresponds to a signal model composed of a smooth component plus a piecewise constant one. Then equation (54) specializes to

$$\dot{x}_\tau(t) = [\dot{x}_\tau](t) + \mu_0 \delta(t - t_r),$$

which reads in operational domain $\mathcal{M}$, as

$$s\hat{x}_\tau - x_\tau(0) - p(s) = \mu_0 e^{-t_r s}, \tag{55}$$

where $p(s)$ is the analogue of $[\dot{x}_\tau](t)$ in $\mathcal{M}$. Applying the differential operator $\Pi = \frac{d}{ds} + t_r$ to both members of equation (55), we get ride of the exponential on the right hand side and obtain:

$$s\hat{x}'_\tau + \hat{x}_\tau - p'(s) = -t_r\{s\hat{x}_\tau - x_\tau(0) - p(s)\} \tag{56}$$

It remains now to annihilate the unknown terms $p(s)$ and $p'(s)$ in order to obtain an estimor for t_r. From the assumption on $[\dot{x}_\tau](t)$, it follows that $s^{\kappa+1}p(s)$ and $s^{\kappa+1}p'(s)$ are polynomials in the variable s, of degree not exceeding κ and $\kappa - 1$ respectively. The differential operator $\frac{d^{\kappa+2}}{ds^{\kappa+2}} \cdot s^{\kappa+1}$ thus eliminates their contribution in (56) as well as that of the initial condition $x_\tau(0)$. This yields

$$\frac{d^{\kappa+2}}{ds^{\kappa+2}} s^{\kappa+2}\hat{x}'_\tau + \frac{d^{\kappa+2}}{ds^{\kappa+2}} s^{\kappa+1}\hat{x}_\tau = -t_r \frac{d^{\kappa+2}}{ds^{\kappa+2}}\{s^{\kappa+2}\hat{x}_\tau\}. \tag{57}$$

Finaly dividing by s^ν, $\nu \geqslant \kappa + 3$ and replacing x_τ by y_τ, leads, in the time domain, to the linear estimator

$$\int_0^T p_{\nu,\kappa}(t) y_\tau(t) dt = -\tilde{t_r}(\tau) \int_0^T q_{\nu,\kappa}(t) y_\tau(t) dt \tag{58}$$

where

$$p_{\nu,\kappa}(t) = (-1)^{\kappa+3} \frac{(T-t)^{\nu-\kappa-3}}{(\nu-\kappa-3)!} t^{\kappa+3} + \sum_{m=0}^{\kappa+1} a_\kappa(m) \frac{(T-t)^{\nu-\kappa-2+m}}{(\nu-\kappa-2+m)!} t^{\kappa+2-m},$$

with

$$a_\kappa(m) = (-1)^{\kappa+2-m} \frac{[(\kappa+1)!]^2}{m![(\kappa+1-m)!]^2} \left\{ \frac{(\kappa+2)^2}{m+1} + \frac{1}{\kappa+2-m} \right\}$$

and

$$q_{\nu,\kappa}(t) = \sum_{m=0}^{\kappa+2} (-1)^{\kappa+2-m} \binom{\kappa+2}{m} \frac{(\kappa+2)!(T-t)^{\nu-\kappa-3+m} t^{\kappa+2-m}}{(\kappa+2-m)!(\nu-\kappa-3+m)!}.$$

Observe that the polynomials $p_{\nu,\kappa}(t)$ and $q_{\nu,\kappa}(t)$ do not depend on τ ; they are computed once for all estimation intervals I_τ^T. In the sequel, the estimator given above in (58) will be named *Estimator A*.

5.2.2 *Estimator B*

The second estimator, subsequently named *Estimator B*, will follow by a simplification in the design of *Estimator A*. This simplification amounts to ignoring the regular part of the signal derivative. A piecewise constant model is thus assumed for the signal and hence, (56) reduces to

$$s\hat{x}'_\tau + \hat{x}_\tau = -t_r\{s\hat{x}_\tau - x_\tau(0)\} \tag{59}$$

Following the same steps below (56), we get in the time domain the following estimator, for each value of $\tau \geq 0$:

$$\int_0^T [(\nu+1)t - 2T](T-t)^{\nu-2}\, t\, y_\tau(t)dt = \widetilde{t_r}(\tau) \int_0^T (\nu t - T)(T-t)^{\nu-2} y_\tau(t)dt. \tag{60}$$

5.2.3 *Estimator C*

Here, we assume a piecewise affine signal model. Equation (54), with $N = 2$ then becomes

$$\ddot{x}_\tau(t) = \mu_1 \delta(t - t_r) + \mu_0 \frac{d}{dt}\delta(t - t_r), \tag{61}$$

Following the same steps as before, we get in the operational domain:

$$s^2\hat{x}_\tau - s x_\tau(0) - \dot{x}_\tau(0) = (\mu_1 + \mu_0 s)e^{-t_r s}.$$

Upon noting that the right member of this equation satisfies the differential equation $u'' + 2t_r u' + t_r^2 u = 0$, we deduce the system of equations

$$-\frac{d^i}{ds^i}(s^2\hat{x}''_\tau + 4s\hat{x}'_\tau + 2\hat{x}_\tau) = 2t_r \frac{d^i}{ds^i}(s^2\hat{x}'_\tau + 2s\hat{x}_\tau) + t_r^2 \frac{d^i}{ds^i}(s^2\hat{x}_\tau); \quad i = 2, 3 \tag{62}$$

Dividing by s^ν, $\nu > 2$ and replacing x_τ by y_τ, leads, in the time domain, to a linear system of equations of the form

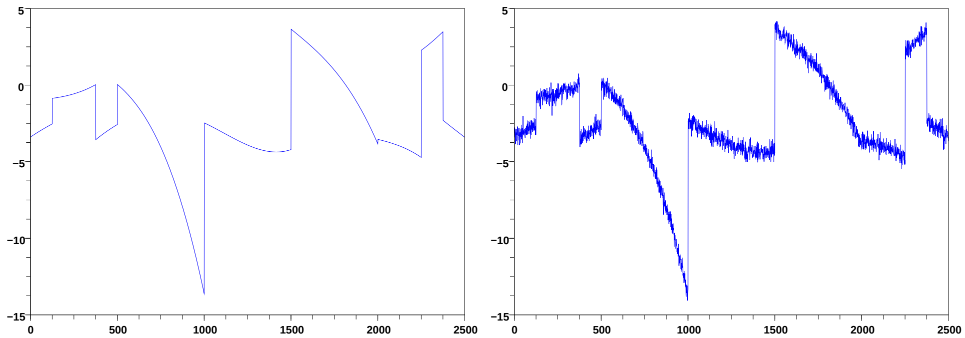
$$\mathcal{P}_\nu(\tau)\widetilde{\mathbf{\Theta}}(\tau) = \mathcal{Q}_\nu(\tau), \tag{63}$$

where the parameter vector is $\widetilde{\mathbf{\Theta}} = [\widetilde{t_r},\, \widetilde{t_r^2}]^t$.

A detailed analysis and comparison of these estimators will be presented in a forthcoming paper. In the sequel, we give some numerical simulation of *Estimator C*.

5.2.4 Numerical simulation

The simulated noise-free signal and a sample realization of its noisy observation $y(t)$ in (53) are depicted in figure 11. Each of the nine segments is a



(a) Piecewise polynomial signal (b) Noisy observation signal, $SNR = 21dB$

Fig. 11. Simulated signals

polynomial of degree 3. The width of each estimation interval I_τ^T is set to $T = 100$ samples and the order of the iterated integrals is $\nu = 5$. Recall that, unless the interval I_τ^T contains a change point, $t_r(\tau)$ must be equal to zero, and so for both $\mathcal{P}_\nu(\tau)$ and $\mathcal{Q}_\nu(\tau)$ (up to the noise level). Therefore, we set $\widetilde{t_r}(\tau) = 0$ for those values of τ for which $\|\mathcal{P}_\nu(\tau)\|_1 < \varrho \max_{\tau\in[0,\,2400]} \|\mathcal{P}_\nu(\tau)\|_1$, for some threshold $\varrho > 0$. The following figure displays $\|\mathcal{P}_\nu(\tau)\|_1$ (top) and $\|\mathcal{Q}_\nu(\tau)\|_1$ (bottom), divided by $\max_{\tau\in[0,\,2400]} \|\mathcal{P}_\nu(\tau)\|_1$, as a function of τ.

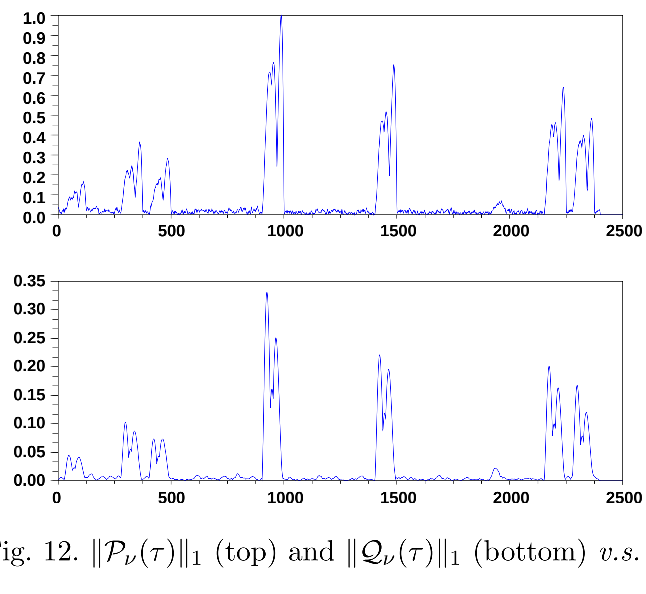


Fig. 12. $\|\mathcal{P}_\nu(\tau)\|_1$ (top) and $\|\mathcal{Q}_\nu(\tau)\|_1$ (bottom) *v.s.* τ.

The estimates $\widetilde{t_r}(\tau)$ and $\widetilde{t_r^2}(\tau)$ obtained with the threshold $\varrho = 0.04$ are shown in figure 13.

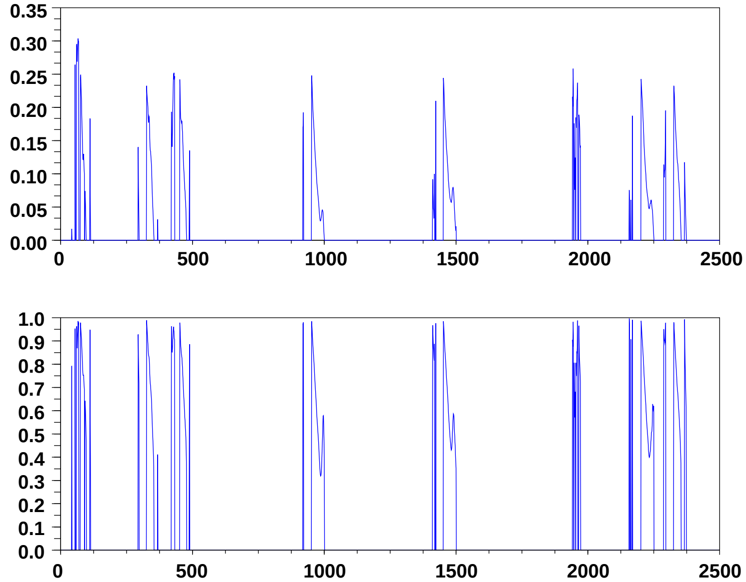


Fig. 13. Change point estimation: $\widetilde{t_r}(\tau)$ (top) and $\widetilde{t_r^2}(\tau)$ (bottom) *v.s.* τ.

Next we present the obtained results over 1000 simulations. These are summarized in figure 14.

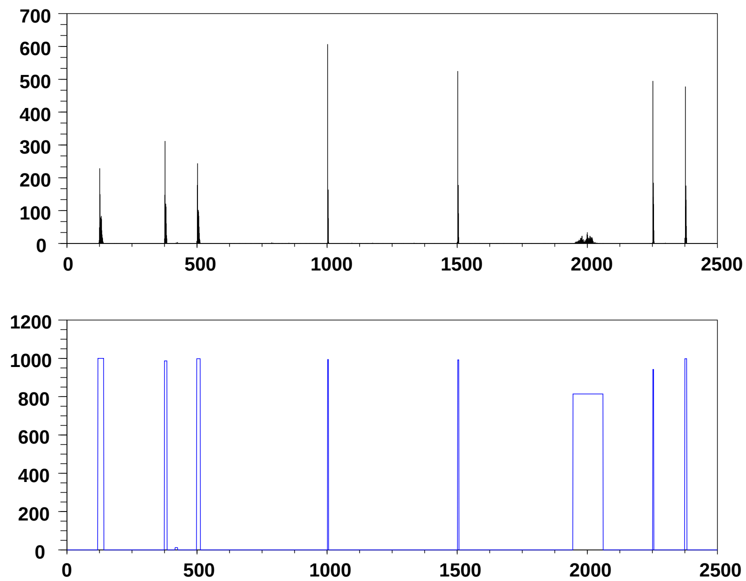


Fig. 14. Distribution of the estimated change point estimation over 1000 trials.

The top plot of figure 14 displays the distribution of the estimated change points. The height of each rectangle in the bottom figure (almost all are centered around an exact change point) indicates the number of estimated change points within it. Even the very small jump, located at $t = 2000$, is reasonably correctly estimated: more than 80% of correct estimation for a location error not exceeding $\pm T/2$, *i.e.* ± 50 samples over 2500. Moreover, these results show a good robustness againts overestimation of the number of change points. This is to be contrasted with estimators based on the popular Akaike information criteria and the Bayesian information criteria which are known to strongly overestimate the number of change points [39].

6 Concluding remarks

We have presented a new standpoint for parametric estimation. This standpoint, recently emerged in control theory, derives from differential algebra, noncommutative ring theory, and operational calculus. We have shown, through several examples and applications, how very simple estimation algorithms with good robustness to noise can be devised within the framework of such unusual mathematical chapters in signal processing. Surprisingly, a least squares interpretation is shown to be attached to the presented approach. It has allowed us to give a first step towards a more complete analysis of the proposed estimation algorithms. Some important questions are however still open. In particular, studies aiming at providing performance measures which fit with the present mathematical framework are in progress, both in signal processing and in control.

References

[1] M. Fliess, H. Sira-Ramírez, An algebraic framework for linear identification, in: ESAIM: COCV, Vol. 9, SMAI, 2003, uRL: http://www.emath.fr/cocv.

[2] M. Fliess, M. Mboup, H. Mounier, H. Sira-Ramírez, Questioning some paradigms of signal processing via concrete examples, in: H. Sira-Ramírez, G. Silva-Navarro (Eds.), Algebraic Methods in Flatness, Signal Processing and State Estimation, Innovación ed. Lagares, México, 2003.

[3] A. Buium, Differential Algebra and Diophantine Geometry, Hermann, Paris, 1994.

[4] E. Kolchin, Differential Algebra and Algebraic Groups, Academic Press, New York, 1973.

[5] M. Fliess, Some basic structural properties of generalized linear system, System and Control Letters 15 (1990) 391–396.

[6] M. Fliess, Une interprétation algébrique de la transformation de laplace et des matrices de transfert, Linear Algebra and its Application (203-204) (1994) 429–442.

[7] M. Fliess, Automatique en temps discret et algèbre aux différences, Forum Math 2 (1990) 212–232.

[8] M. Fliess, J. Lévine, P. Martin, P. Rouchon, Flatness and defect of non-linear systems: introductory theory and applications, International Journal of Control 61 (1995) 1327–1361.

[9] H. Sira-Ramírez, S. Agrawal, Differentially flat systems, Marcel Dekker, New York, 2004.

[10] J. Mikusiǹski, Operational Calculus, Vol. 1, PWN Varsovie & Oxford University Press, Oxford, 1983.

[11] J. Mikusiǹski, T. K. Boehme, Operational Calculus, Vol. 2, PWN Varsovie & Oxford University Press, Oxford, 1987.

[12] K. Yosida, Operational Calculus - A Theory of Hyperfunctions, Springer, New York, 1984.

[13] M. Basseville, I. V. Nikiforov, Detection of Abrupt Changes: Theory and Application, Prentice-Hall, 1993.

[14] R. Kumaresan, D. W. Tufts, Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise, IEEE Trans. Acoust., Speech, Signal Processing 30 (6) (1982) 833–840.

[15] Y. Bresler, A. Macovski, Exact maximum likelihood parameter estimation of superimposed exponential signals in noise, IEEE Trans. Acoust., Speech, Signal Processing 34 (10) (1986) 1081–1089.

[16] M. A. Rahman, K. B. Yu, Total least squares approach for frequency estimation using linear prediction, IEEE Trans. Acoust., Speech, Signal Processing 35 (10) (1987) 1440–1454.

[17] M. Kahn, M. Mackisack, M. Osborne, G. K. Smyth, On the consistency of Prony's method and related algorithms, J. Comput. Graph. Statist. 1 (1992) 329–349.

[18] Y. Li, K. Liu, J. Razavilar, A parameter estimation scheme for damped sinusoidal signals based on low-rank hankel approximation, IEEE Trans. Signal Processing 45 (2) (1997) 481–486.

[19] E. Djermoune, M. Tomczak, Statistical analysis of the kumaresan-Tufts and matrix pencil methods in estimating a damped sinusoid, in: EUSIPCO'04, Vienna, Austria, 2004.

[20] J. R. Traper, H. Sira-Ramírez, V. F. Batlle, An algebraic frequency estimator for a biased and noisy sinusoidal signal, Signal Processing 87 (6) (2007) 1188–1201.

[21] H. Sira-Ramìrez, M. Fliess, On discrete-time uncertain visual based control of planar manipulators: an on-line algebraic identification approach, in: Proc. IEEE CDC, Las Vegas, 2002.

[22] S. Fuchshumer, Algebraic linear identification, modelling, and applications of flatness-based control, Ph.D. thesis, Johannes Kepler Universität, Linz (2006).

[23] M. Fliess, S. Fuchshumer, K.Schlacher, H. Sira-Ramirez, Discrete-time linear parametric identification: An algebraic approach, in: JIME'2006, Poitiers, France, 2006.

[24] K. Perlin, An image synthesizer, in: SIGGRAPH'85, 1985, pp. 287–296.

[25] T. Mikosch, Elementary stochastic calculus with finance in view, Vol. 6 of Advanced Series on statistical science and applied probability, World scientific, 1998.

[26] M. Atiyah, I. Macdonald, Introduction to Commutative Algebra, Addison-Wesley, MA, 1969.

[27] J. McConnell, J. Robson, Noncommutative Noetherian Rings, Amer. Math. Soc., Providence, RI, 2000.

[28] K. Yosida, Functional Analysis, 4th Edition, Springer-Verlag, New York, 1974.

[29] S. Diop, M. Fliess, Nonlinear observability, identifiability and persistent trajectories, in: Proc. 36^{th} IEEE Conf. Decision Control, Brighton, 1991, pp. 714–719.

[30] M. Fliess, H. Mounier, A trajectory interpretation of torsion-free controllability, IMA J. Math. Control Informat. 10 (2002) 49–57.

[31] M. Fliess, Analyse non standard du bruit, CRAS, Série 1, Mathématiques 342 (2006) 797–802.

[32] N. J. Higham, A survey of componentwise perturbation theory in numerical linear algebra, in: W. Gautschi (Ed.), Mathematics of Computation 1943–1993: A Half Century of Computational Mathematics, Vol. 48, AMS, Providence, RI, USA, 1994, pp. 49–77.

[33] R. D. Skeel, Scaling for numerical stability in Gaussian elimination, J. Assoc. Compt. Mach. 26 (1979) 494–526.

[34] V. Frayssé, The power of backward error analysis, Ph.D. thesis, Inst. Nat. Polytechnique de Toulouse, CERFACS, hDR (2000).

[35] S. Saha, S. Kay, Maximum likelihood parameter estimation of superiroposed chirps using monte carlo importance sampling, IEEE Trans. Signal Processing 50 (2002) 224–230.

[36] B. Volcker, B. Ottersten, Chirp parameter estimation from a sample covariance matrix, IEEE Trans. Signal Processing 49 (3) (2001) 603–612.

[37] M. Morvidone, B. Torresani, Time scale approach for chirp detection, Int. J. Wavelets Multiresolut. Inf. Process. 1 (1) (2003) 19–50.

[38] I. Gijbels, P. Hall, A. Kneip, On the estimation of jump points in smooth curves, Ann. Inst. Statist. Math. 51 (1999) 231–251.

[39] M. Lavielle, Using penalized contrasts for change-point problem, Signal Processing 85 (2005) 1501–1510.

[40] S. Mallat, A Wavelet Tour of Signal Processing, 2nd Edition, Academic Press, 1999.

[41] M. Raimondo, N. Tajvidi, A peaks over threshhold model for change-point detection by wavelets, Statistica Sinica 14 (2004) 395–412.

[42] M. Mboup, C. Join, M. Fliess, A revised look at numerical differentiation with an application to nonlinear feedback control, Tech. rep., Submitted to MED'07 (2007).

[43] L. Schwartz, Théorie des distributions, 3rd Edition, Hermann, 1998.